# Standard Evaluation Procedure

# for the

# Multigeneration Reproduction Study

Draft
November 2, 1992

Prepared by the following:

Gary Burin, Ph.D., DABT
Laurence Chitlik, M.S., DABT
James Rowe, Ph.D.
Susan Makris, Ph.D.

The significant contributions of Dr. Quang Bui to early work on this document are gratefully acknowledged by the authors.

TABLE OF CONTENTS

## I. Introduction

The multigeneration reproduction study determines potential adverse effects of a treatment regimen to the male and female reproductive systems, the conceptus and to the neonate. It should be regarded as a screening assay for a wide variety of endpoints related to reproduction. The multigeneration reproduction study which is properly designed, conducted and interpreted can detect effects on germ cells, gametogenesis, libido, fertilization, implantation, embryonic, fetal, neonatal growth and development, lactation, and postweaning growth and maturity. The multigeneration reproduction study also provides information regarding the direct toxicity of a chemical to the pregnant animal. Due to the profound physiological changes which occur during pregnancy, these effects may be much different than those observed in chronic and subchronic studies.

The Agency guidelines for reproductive toxicity risk assessment are intended to ensure a consistent approach to evaluations. Interpretation of endpoints related to reproductive toxicity should be viewed in the context of these guidelines. The most relevant guidelines on this topic are the (1) Guidelines for Male Reproductive Risk Assessment (USEPA, 1988) and the (2) Guidelines for the Health Assessment of Suspect Female Reproductive Toxicants USEPA, 1988). **[The new Agency guidelines will be cited here after finalization]** Familiarity with these Agency risk assessment guidelines will provide the framework for the evaluation of the multigeneration reproduction study.

Special studies such as the continuous breeding protocol and studies routinely required in OPP such as the chronic feeding study and developmental toxicity studies provide additional information which may be relevant to reproductive toxicity. Although the focus of this Standard Evaluation Procedure (SEP) is on the

multigeneration reproduction study, aspects of other studies relevant to reproductive toxicity will be discussed here as well. In addition, because reproductive toxicants may not cause effects solely by affecting the integrity of reproductive tissues but may also induce genetic abnormalities, adverse developmental effects and other toxic effects, the reviewer is referred to other relevant guidelines such as the Guidelines for Mutagenicity Risk Assessment (USEPA, 1986), the Guidelines for Carcinogen Risk Assessment (USEPA, 1986), and the Guidelines for Suspect Developmental Toxicants (USEPA, 1991).

**A. Definitions**

For the purpose of this Standard Evaluation Procedure (SEP), the following definitions are noted. Some of these terms are presented for explanatory purposes only as they are not often used in the routine evaluation of the multigeneration reproduction study.

1. Reproductive toxicity - The occurrence of adverse effects on the reproductive system that may result from exposure to environmental agents. The toxicity may be expressed as alterations to the female or male reproductive organs or the related endocrine system. The manifestation of such toxicity may include but not be limited to, adverse effects on onset of puberty, gamete production and transport, cyclicity, sexual behavior, fertility, gestation, parturition, lactation, pregnancy outcomes, premature reproductive senescence, or modifications in other functions that are dependent on the integrity of the reproductive system.

2. Female reproductive cycle - The periodic recurrence of events in the neuroendocrine and generative systems (hypothalamus, pituitary, uterus, ovaries, and accessory sexual structures) associated with estrus in lower mammals and menstruation in humans and nonhuman primates.

3. Fertility - Ability to conceive and to produce offspring within a defined period of time. For litter-bearing species,

fecundity (the number of offspring) is also a measure of fertility.

a. Fertile - Having a level of fertility that is within or exceeds the normal range for that species.

b. Subfertile - Having a level of fertility that is below the normal range for that species but not infertile.

c. Infertile - Lacking fertility for a specified period of time. The infertile condition may be temporary; permanent infertility is termed sterility.

4. Male reproductive system - Those processes and organs in the male that are involved directly in sexual behavior and procreation. For this document, these include the testes, epididymides, vas deferens, accessory sex glands, penis, pituitary and hypothalamus. Not all regions of the latter two organs are considered to have a role in reproductive function.

5. Male reproductive toxicity - The occurrence of adverse effects on the male reproductive system that may result from exposure to some agents. The toxicity may be expressed as alterations to the male reproductive organs and/or the related endocrine system. The manifestations of such toxicity may include alteration in sexual behavior, fertility, pregnancy outcomes, or modifications in other functions that are dependent on the integrity of the male reproductive system.

**B. When Required**

The Office of Pesticide Programs requires a multigeneration study when the use of the pesticide may result in dietary exposure to the pesticide, its active ingredient(s), metabolite(s) or degradation product(s) (i.e. when tolerances or exemptions from tolerances are considered) and, for nonfood uses, if exposure is expected over a portion of the human lifespan "which is significant in terms of the frequency of exposure, magnitude of exposure or the duration of exposure" or reproductive concerns arise from other studies. The establishment of a temporary tolerance for residues in food may require a multigeneration study depending upon factors

such as extent and duration of exposure, structure-activity concerns and the results of subchronic, developmental and mutagenicity studies. Part 158 (40 CFR) states that at least an interim report from the first generation of multigeneration reproduction study is needed if the theoretical maximum residue contribution is greater than 50% of the maximum permitted intake.

Structural similarities to reproductive toxicants may lead to the need for reproductive toxicity testing when it otherwise would not be required for nonfood uses. However, small differences in structure may lead to major differences in the potential for reproductive toxicity. Screening assays for reproductive toxicity (such as the Chernoff-Kavlock screen or the reproduction screen included in the OECD Screening Information Data Sets (SIDS) testing battery) are generally not requested in lieu of a multigeneration study but may trigger the need for further testing. As noted elsewhere in this SEP, limited information regarding reproductive toxicity is available from subchronic, chronic and developmental toxicity studies and this information may also be considered in making the determination of whether to require a multigeneration reproduction study.

**C. Utility of the Reproduction Study in a Regulatory Setting**

The data obtained from the two-generation reproduction study provide information concerning the potential systemic and reproductive effects of agents resulting from <u>in utero</u> through adult exposure. Such information may be utilized to set limits regarding the dietary intake of a pesticide i.e. the Reference Dose (RfD) or may be extrapolated to assess the risk to workers or bystanders exposed to the pesticide (i.e. the Margin of Exposure). Reproductive toxicity concerns are the endpoint upon which the RfD has been established for a growing number of pesticides and has been a trigger for Special Review for several chemicals. The multigeneration study may also lead to additional testing in

laboratory animals or epidemiological observations in man.

The reproduction study should be regarded as a screening assay which examines a wide variety of reproductive endpoints which are of potential concern. Unfortunately, however, the design does not readily permit a determination as to whether the effects are male or female-mediated. Further investigations or modifications of the routine protocol would be required for such determinations. Thus, when an effect on reproduction is found in the multigeneration reproduction study, follow-up studies may be necessary to elucidate the source of the effect and to examine, in more specific and sensitive ways, the endpoint that appears to be effected. It may be necessary for regulatory purposes to identify the gender in which the effects arise or to determine the effects of a chemical upon sperm parameters which normally have not been routinely assayed in the multigeneration study. In other cases, effects may be observed in the multigeneration study which could arise from multiple causes but which may not need further investigation for the purpose of hazard identification. For example, decreased neonatal weight gain, which occurs at dose levels similar to that which induces toxicity in the adult, may result from direct exposure of the neonate to the test compound during lactation, effects induced <u>in utero</u>, or effects upon the lactational ability or behavior of the dam. If this study is the basis for the RfD, a cross-fostering study may help to determine whether the decreased pup weight gain is an effect induced during lactation or whether it is the delayed manifestation of developmental toxicity induced <u>in utero</u>. If this effect is not the basis for the establishment of an RfD and it has been adequately investigated in the multigeneration study for the purpose of establishing both effect level(s) and NOELs, further studies may not be needed.

It is recognized that certain toxic manifestation that cause frank systemic toxicity in animals can affect reproductive performance. Reductions in food consumption, body weight gain and

food efficiency can concomitantly have associated effects upon pup weight and/or litter size, lactation indices and other end-points evaluated in a reproduction study. Therefore, it would be incorrect to assume that the pesticide being tested is a "selective" reproductive toxicant since all of the manifestations observed are obviously interrelated.

Although the primary intent of the study is to assess potential reproductive effects, it is often impossible to separate these effects from other manifestations of general/systemic toxicity within the study. Since the separation of effects into "primary reproductive effects" or as effects "potentially secondary to other toxic manifestations" is sometimes difficult, the establishment of a NOEL for reproductive effects should be undertaken only when it is clear that there is selective reproductive toxicity. This may be determined in either of two ways. The first would be by the finding of a lower NOEL for alterations in reproductive tissues or parameters related to reproductive ability than for other manifestations of toxicity. The second manner in which this could be determined would be by a knowledge of the mechanism of action of a compound. For example, chemicals which interfere with microtubule formation, such as benomyl and its carbendizim metabolite, clearly are toxic to rapidly dividing cells such as sperm (Gray et al, 1990). Although effects are seen at other sites as well, a preferential sensitivity to the male reproductive system can be explained based upon knowledge of the mechanism of toxicity.

The severity of reproductive toxicity compared to more general toxicity is also an indication of selective reproductive toxicity. For example, a slight depression in parental body weight gain has little effect on reproductive parameters. Additional data obtained with a careful spacing of dose levels may be necessary to differentiate reproductive toxicity from other more general forms of toxicity.

In the past, reproduction studies have had several types of NOEL's assigned to them, e.g., "Reproduction Toxicity NOEL's" and "Systemic Toxicity NOEL's". Some reviewers have even assigned "Developmental Toxicity NOEL's" to end-points observed in reproduction studies. The fact is, however, that in a substantial percentage of the reproduction studies, the distinction between these various NOEL's is neither possible nor appropriate. Therefore, this SEP will depart from this established practice and generally require that the reviewer establish only one NOEL. This NOEL will be based on the absence of any toxic manifestations in the reproduction study.

In a situation where the reviewer considers the effect to be selective to the reproductive system, the reviewer should examine other available toxicity studies in the data base before finalizing this assessment. Chronic and subchronic studies should be carefully examined in order to evaluate the target toxicity of the pesticide. Only after this assessment is completed can the reviewer make any determination as to the necessity for a separate "Reproductive Toxicity NOEL".

**D. Correlation with other Relevant Data**

The multigeneration study should be evaluated in the context of all other relevant information including developmental toxicity studies, subchronic and chronic studies, and metabolic and pharmacokinetic data. These studies are often conducted in the same species (the rat) as is the multigeneration study and at similar dose levels. Organ weights and histopathology of the ovary, uterus, testes, epididymides, seminal vesicles, prostate gland, pituitary and hypothalamus in chronic and subchronic studies may provide indications of target organs associated with reproductive toxicity. This information should be carefully considered in conjunction with the multigeneration study findings.

Data regarding other endpoints (e.g. organ weight and histopathology from nonreproductive organs, clinical observations and body weight data) are useful in understanding the general toxicity of the test compound. As noted above, comparison of dose levels inducing reproductive and nonreproductive toxicity and an understanding of the association between the various manifestations of toxicity may allow a determination of whether a chemical is a selective reproductive toxicant rather than a chemical which has reproductive effects secondary to other toxic manifestations.

Testicular histopathology data from the subchronic and chronic studies should be compared carefully to the results of the multigeneration study. Formalin fixation combined with paraffin embedding of the testis may result in artifacts such as shrinkage, vacuoles and clumping of nuclear material which can mask effects and impair meaningful interpretation. The lack of an effect on the testis may simply reflect poor specimen preparation. A careful description of testicular histology facilitates study interpretation and is necessary for the separation of artifacts from compound-induced effects. Cell staging or morphometric measurements also facilitate the interpretation of testicular histology. If histopathological changes in the testes are observed in the subchronic or chronic studies, effects on fertility may or may not be observed at similar dose levels in the multigeneration study. The lack of sensitivity of the multigeneration study for the detection of effects on fertility limits the ability of the study to confirm reproductive toxicity indicated by histological changes in reproductive tissues.

Although multigeneration reproduction and developmental toxicity studies examine some of the same endpoints, the results of these studies should also be compared cautiously. The filial or second generation and the developmental toxicity study both involve <u>in utero</u> exposure to the embryo and fetus. However, there are several important differences between the two studies. Among the

most important of these differences are the patterns of exposure in the two studies. First, the dose levels in the developmental toxicity study are usually much higher than those which are utilized in the multigeneration study. Second, the period of treatment in the developmental toxicity study is short, usually encompassing only the period of major organogenesis. Dosing may occur during a critical period in the developmental toxicity study, and the short duration of exposure may not result in the achievement of a steady state in the fetal/embryonic compartments at the critical time of gestation. The third major difference related to exposure is that the route of administration is usually oral (gavage) in the developmental toxicity study versus dietary exposure in the reproduction study. This may result in much higher peak plasma levels of the test material due to more rapid absorption in the study conducted by gavage. The bioavailability of the test material may also be greater via the gavage route of administration than after incorporation in the diet, resulting in a greater area under the plasma curve (AUC).

There are other differences between the studies which are not related to exposure. The level of individual fetal examination is generally more extensive in the developmental toxicity study while behavioral effects resulting from <u>in utero</u> exposure may be more amenable to study in the reproduction assay. Cannibalism of pups with abnormalities may reduce the sensitivity of the multigeneration study in the detection of malformations. Because the fetuses are delivered by caesarian section, cannibalism will not be a factor in the developmental toxicity study. Given the above, it is not surprising that qualitative and quantitative differences are often observed between multigeneration reproduction studies and developmental toxicity studies.

Information available in humans is potentially the most important ancillary information to be utilized in the assessment of the potential of a chemical to cause reproductive toxicity, since

the objective in the review of toxicity studies is the protection of human health. Clusters or case reports of reproductive toxicity should be considered in light of the toxicity observed in multigeneration studies. Although well conducted epidemiology studies pertaining to reproductive toxicity are rarely available for individual pesticides, such studies have provided important confirmatory evidence of male reproductive toxicity for ethylene dibromide and dibromochloropropane (Whorton and Milbey., 1980). The development of appropriate biomarkers for reproductive toxicity may facilitate future investigations of reproductive toxicity in humans and allow a better correlation of effects observed in animal studies with those observed or predicted in humans (See Table 10 in Risk Assessment Section VI).

Comparisons of parameters such as reproductive indices can also be made between litters and between generations in a multi-generation study. However, in the design of a multigeneration reproduction study, the length of exposure between parental (P generation) and filial (F generation) is not equal. Treatment of the P generation commences with adult animals (at least 6 weeks of age) whereas filial (F1 and F2) generations are exposed continuously <u>in utero</u> from conception, throughout gestation and lactation prior to evaluation. The F1 generation is also exposed and observed through puberty and into adulthood. Therefore, adverse reproductive effects observed in any particular generation are not necessarily indicators or predictors of similar adverse effects in another generation. Similarly, successive litters of the same generation (Fa and Fb) should not be treated as replicates. Chemicals which bioaccumulate (have long half-lives for elimination and thus require prolonged exposure to reach steady state) may have a greater incidence or severity of effects observed in the second litter (F1b or F2b) as compared with the first litter (F1a and F2a) (Christian, 1986). Parity, age and body weight are also different at the time of the second mating.

## II. Data Acceptability

### A. General Requirements

**[Add acceptance criteria]**

The US EPA Office of Pesticide Programs, Pesticide Assessment Guidelines "Subdivision F: Hazard Identification: Humans and Domestic Animals", Sections 80-4 and 83-4 and the OECD test guideline 416 (May 26, 1983) define protocol and reporting requirements for multigeneration reproduction studies. In all cases, scientific judgment must be exercised regarding the deviations from published guidelines. The guidelines should not be construed as absolute requirements and modifications of standards protocols may be appropriate on a case-by-case basis. Significant deviations from the prescribed guidelines require an adequate justification from the testing laboratory. Experts within the Health Effects Division should be consulted if there is any uncertainty regarding study acceptability.

Some of common issues concerning study acceptability are outlined below:

1. Individual animal data must always be provided in submitted study reports. These data should allow for tracing of each parent and its offspring throughout the two generation study. Without such data, it is not possible to determine whether sibling matings have occurred.

2. The highest dose level must induce toxicity (either general or reproductive). This is necessary to maximize the sensitivity of the study. It is preferable that the test chemical not induce mortality in treated animals. The lowest dose should not induce any effects in parents or offspring.

3. Although standardization (culling) of litters is suggested in the US EPA Guidelines, it is not a requirement for acceptability. Standardization facilitates study conduct and statistical analysis but may also reduce study sensitivity, resulting in the loss of information. Studies conducted under the

recommendations of the OECD guidelines for reproduction studies will generally not include standardization in the study design, but such a protocol would not jeopardize study adequacy.

4. A one-generation reproduction study is not acceptable to fulfill the data requirement for a multigeneration study. Effects are sometimes observed only in the second generation. See Section IV.A.

5. Compound administration must be continuous throughout the study. It is preferable that dosing remain constant on a body weight basis throughout the premating, mating, lactation and weaning periods.

6. The study must include an adequate histopathological investigation of the reproductive tissues since changes in fertility and other reproductive indices are often observed only at dose levels greater than those inducing histopathological changes. See discussion under "Evaluation of Study Conduct".

7. Studies should comply with Good Laboratory Practice requirements established by either EPA or OECD.

**B. Acceptability of Other Protocols**

Some basic comparisons between the EPA study design and other commonly followed protocols are shown in the Table 1.

The differences between these protocols are relatively minor and should not present a barrier to study acceptance. Efforts are currently under way to develop a harmonized guideline which is acceptable to EPA (under both FIFRA and TSCA) and to OECD. Until then, Agency policy directs acceptance of study protocols which were designed according to OECD recommendations.

The FDA multigeneration study protocol referenced in Table 1 is described in more detail below. In addition, FDA Segments I, II, and III protocols and Reproductive Assessment by Continuous Breeding (RACB) are discussed.

Table 1. Comparison between multigeneration study designs recommended by US and foreign regulatory agencies

| Protocol by Agency[a] | EPA[b] | FDA | OECD |
|---|---|---|---|
| No./sex/group | 20 | 30 for P<br>25 for F1 | 20 |
| Age at start of dosing (P) | 8 weeks | After weaning | 5 to 9 weeks |
| Premating period (P) | 10 weeks | 10 weeks | 10 weeks |
| Mating ratio (M:F) | 1:1 | 1:1 | 1:1 or 1:2 |
| Standardization (culling) | Yes[c] (4 males, 4 females) | Yes (5 males, 5 females) | Optional (4 males, 4 females) |
| Disposition of P animals | P males after mating, females after weaning | P males after mating, females after weaning | P males after mating, females after weaning |
| Parental selection (F1) | 1 male and 1 female/litter; no sibling matings | 2 males & 2 females/litter | 1 male & 1 female/litter; no sibling matings |
| F1 premating period | 14 weeks postweaning | 13 weeks postweaning | 10 weeks postweaning |
| Age at start of dosing | 8 weeks | After weaning | 5 to 9 weeks |

a Criteria listed are for rats.
b EPA FIFRA guidelines only.
c Interpreted as optional.

1. FDA Multigeneration Study

The initial three generation reproduction study was developed by the US Food and Drug Administration for the testing of intentional food additives (food coloring agents, sweeteners, etc.) and unintentional food additives including pesticides (Lehman, 1949). The protocol was modified in the early 1960's and was recommended by the President's Science Advisory Committee (1963) for the testing of pesticides. It has been retained with only

slight modification as the multigeneration reproduction study which is required by the Office of Pesticide Programs. The original protocol called for treatment to be initiated in mature male rats 60 to 80 days prior to mating while females were to be treated for 14 days prior to mating. Daily vaginal inspections would be made with the finding of sperm or a copulation plug considered as day 0 of pregnancy. Each generation was required to produce two litters with the first litter discarded at weaning. Two males were originally paired with each female (this was later revised to a 1:1 mating ratio) and extra young and adults were sacrificed. After the young from the second mating of the final generation are weaned, histopathological examinations and organ weight analyses of 10 male and 10 female offspring per group were performed.

2. The FDA Segment I, II, and III Studies

The FDA Segment I, II and III screen (Goldenthal, 1966) was originally proposed to replace the two-litter rat reproduction study for the appraisal of new drugs for use during pregnancy and in women of childbearing potential. These segments are: I. Study of Fertility and General Reproductive Performance; II. Teratological Study; and III. Perinatal and Postnatal Study. The Segment I, II, and III studies, either alone or in combination, cannot be used to satisfy EPA regulatory requirements for a multigeneration study.

a. Segment I

To examine the effects of a drug on male fertility in a Segment I study, male rats should have attained a minimum age of 40 days before drug administration begins. They are treated (via the same route planned for clinical administration) for 60 to 80 days prior to mating to assess effects upon spermatogenesis. These males are then mated with either treated or non-treated females. At a minimum, 10 male animals are mated to 20 females per dose

level.

Adult animals are used to assess toxicological effects upon female fertility. After 14 days of drug administration, the females are mated. Copulation is verified by inspections for sperm or the presence of a copulatory plug which is considered day 0 of pregnancy. At least two dosage levels should be used. The high dose should be the maximum tolerated dosage based upon a slight decrease in body weight gain (or other minimal indications of toxicity) but should not adversely affect the dam with anorexia, sedation or other pharmacological effects.

One half of the females are sacrificed on day 13 and examined for the number and distribution of embryos in each uterine horn, presence of empty implantation sites, and resorptions. The remaining dams are continued on treatment and allowed to deliver normally. The gestation length is determined and the litters examined after delivery for litter size, stillborn and live pups, and gross anomalies. Pups are weighed individually at delivery. Dead pups are preserved for skeletal examination. Pups are weighed on days 4 and 21. At this point it is determined whether a second litter should be initiated or in special cases whether an additional study on reproductive performance of offspring is necessary (matings for additional generations). Thus, the Segment I study is essentially a one-generation reproduction study and cannot be used to satisfy EPA regulatory requirements for a multigeneration study.

b. Segment II

Segment II studies investigate the effects of a chemical on <u>in utero</u> development. They are equivalent to the protocols for the rodent developmental toxicity study required under FIFRA.

c. Segment III

The Segment III perinatal and postnatal study is used to evaluate the effects of drugs on late fetal development, labor and delivery, lactation, and newborn viability and growth (Collins, 1978). It begins with drug administration during the final one-third of gestation and continues through lactation to weaning. Observations are made for effects on parameters such as labor and delivery, the duration of gestation, litter size and pup weight. Continuous compound administration through lactation allows for detection of adverse effects on lactation ability as well as any toxic reactions due to the drug or metabolites on the newborn as a result of excretion in the milk. Some of the offspring may continue to be observed to adulthood.

3. The Continuous Breeding Protocol

The continuous breeding protocol was developed by the National Toxicology Program as a possible alternative to the multigeneration study and has been evaluated using both mice and rat models (Lamb, 1985; Morrissey et al., 1989). The study allows for continuous breeding of treated males and females for 14 weeks (after pretreating for 7 days) with immediate removal of each litter after birth. The exception to this is that the final litter may remain with the dam until weaning or may even remain on study to maturity with subsequent mating. Mating of treated males with untreated females, or treated females with untreated males, can allow identification of gender-specific effects.

This protocol allows for up to five litters per pair during the 14 week study. It also may allow for determination of the affected sex by mating treated animals with untreated controls (crossovers) for 7 days (with treatment discontinued for this period). This protocol offers considerable advantage in that it allows for determination of subfertility due to larger numbers of pregnancies as well as better assessment opportunities for time of onset of effects on fertility. The continuous mating could be

incorporated into the multigeneration study. The second generation (F2) may be taken from the first mating, rather than the last, further shortening the length of the study. Alternatively, the F1 offspring could be used for the continuous breeding segment. This protocol may be acceptable on a case-by-case basis and registrants should be encouraged to discuss individual study designs based on the continuous breeding protocol with scientists prior to study initiation.

4. Combined Reproduction and Developmental Toxicity Study

A modified multigeneration study which includes a developmental toxicity (teratology) phase has occasionally been conducted for pesticides, usually with fetal visceral and skeletal examinations on the second litter from the second (or third) generation. Although this study design may be acceptable for the purpose of assessing reproductive toxicity, it is rarely an adequate assessment of developmental toxicity potential. The dietary exposure levels of the reproduction study are generally inadequate to achieve the degree of maternal toxicity required in a developmental toxicity study. The combined reproduction/ developmental toxicity study will not be discussed further in this SEP.

## III. Evaluation of Study Conduct

### 1. CHOICE OF TEST COMPOUND

Section 83-4 of the 1982 FIFRA Guidelines states that a multigeneration test shall be performed with the technical grade of each active ingredient in the product. The test compound selected should be the technical product intended for commercial use. In some cases testing is conducted using a material which is produced before the commercial manufacture of the technical material, i.e., before manufacturing processes are "on-line". The specifications

of the test material should be available, and the concentration of the active ingredient(s) and of the impurities should be clearly indicated in the study report. This information should also be included in the study evaluation for comparison with material utilized in future studies. If a vehicle is used, it should not produce any systemic or reproductive toxicity. If there is a question as to the toxicity of the vehicle, the registrant should be required to justify the choice of the vehicle.

2. ANIMAL SELECTION

a. Species selection

Criteria in the selection of the test species in a multi-generation reproduction study include animal size, length of gestation, litter size, fertility rate, ovulation, ease of maintenance, and comparability of metabolism of the chemical with that in human (Collins, 1978). Of the species available, the rat and mouse are preferred due to their small size, short gestation time, high fertility rate, spontaneous ovulation, short estrous cycle, and ease of maintenance. Golden Syrian hamsters, although having a short gestation time (15 days) and a large litter size (10-14 pups), are not often used in a reproduction study due to their insufficient and inaccurate fertility rate. Mongolian Gerbils have proven to be useful in a reproduction study (Robinson, 1979) due to their size, short gestation period, ease of maintenance, spontaneous ovulation, and fertility rate. However, gerbils are less prolific than rats or mice, since their litter size is 4-6 pups. In addition, they tend to be monogamous in captivity (Holmes, 1985), but this may be an artifact of housing techniques (Thiessen and Yahr, 1977). Rabbits are rarely used in a reproduction study since they do not ovulate spontaneously and have a relatively long gestation period (30 days). Further, rabbit maintenance has been proven to be expensive. A long gestation period (64-68 days), coupled with a small litter size (1-3 pups)

and a requirement of exogenous source of vitamin C, makes guinea pigs unsuitable for reproductive studies. Primates, dogs, and pigs generally do not provide any major advantage over rats or mice and are not recommended for routine testing.

It has often been stated that the choice of the test animals should be based on similarities to humans with respect to plasma concentration, placental transfer, metabolic patterns, embryonic developmental schedules. However, there is no species which closely resembles humans for all of these parameters. Therefore, for reasons of convenience and the availability of an extensive historical data base, the rat is the preferred species for a reproduction study. An exception is the Fischer 344 strain of rat, which has proven to be an unreliable model for reproduction studies.

b. Health Status

Virgin animals should be selected. Upon arrival at the testing facility, all animals must be quarantined for at least 14 days, during which time their health status must be checked by a veterinarian. Quarantine of newly arrived rodents reduces the possibility of transmitting active infections and allows animals to adjust to their new surroundings.

At study initiation, all animals must be sexually mature (approximately 8-10 weeks old) and disease-free, since treatment or infections during the course of the investigation may lead to unpredictable pharmacological or toxicological outcomes.

c. Number of animals

Reproduction studies customarily require 20 pregnant female rats (or mice) per dosage level; therefore, starting the study with exactly 20 females per group risks study rejection if an inadequate

number of litters is obtained in any group. At least 20 pregnant females at or near term are needed as parental animals for each dose group and numbers fewer than this reduce study sensitivity. The investigators must ensure that an adequate number of litters were obtained for each group and this may have required mating 30 or more females to achieve a sufficient number of litters at term.

Studies with fewer animals per dose level may also be considered as acceptable (core minimum) on a case-by-case basis. Studies with less than 15 litters at any single dose level (with the exception of the high dose level) are generally considered to be inadequate for the purpose of meeting regulatory requirements. If less than 15 litters are available at the high dose level due to compound-related toxicity but adequate numbers are available at other dose levels to establish a NOEL, the study may still be considered as acceptable. Studies which may appropriately be viewed in combination with one or more other reproduction studies may also be adequate for regulatory purposes.

3. ENVIRONMENTAL FACTORS

Environmental factors have an important influence on the toxicity of test compound. Environmental factors which may influence the reproductive performance of the test animals or exert unnecessary maternal stress include the following:

a. Temperature and Humidity

Adaptations to changes in ambient temperature in rodents consist of peripheral vascular constriction, piloerection, increased metabolic activity which may result in increased food consumption, and variations in biologic and physiologic effects in response to fluctuations in the ambient temperature. Daily fluctuations in room temperature and humidity may act as significant stressors (Kohn and Barthold, 1984). Yagil et al.

(1976) reported that the production of milk was impaired in rats exposed to 35°C for 8 hours/daily. Temperature as low as 32°C has also been demonstrated to impair the reproductive capacity of rats (Yamauchi et al., 1981). Room temperatures between 72° and 76°F (22-24°C) are desirable, and the humidity should range between 40 and 60%. It should be noted that high ambient temperature and humidity may increase the susceptibility of the test animals to infectious agents (Baetjer, 1985) and may cause male infertility (Baker, 1979).

b. Light Intensity and Timing

Seventy five - 125 fc (foot-candle) are suggested as an optimal range for light intensity which should be evenly distributed to all animals in the room. However, there is evidence of retinal degeneration in laboratory animals at that recommended light intensity range (Anver and Cohen, 1979; Bellhorn, 1980). Light is a stimulant and synchronizer of the reproductive system (Pakes et al., 1984) and is thought to be mediated through the hypothalamus (Wiehe, 1976). Consequently, photoperiod (light-dark cycle) not only can modify the biologic response but also has a profound effect on the circadian rhythms of rodents (Hastings and Menaker, 1976). For a reproduction study, a daily light period of 12-14 hours appears to be optimum (Mulder, 1971). Standardization of light intensity and duration is necessary in a reproduction study since the cyclicity of estrus and ovulation in rodents is controlled by the diurnal rhythm of the photoperiod. Constant light for as few as 3 days may induce persistent estrus and polycystic ovaries (Baker et al., 1979).

c. Nutrient Requirements

Evaluation of the diet which the animals receive may be neglected in a reproduction study. The final report must contain information concerning diet analysis as well as identification of

contaminants. Nutritionally adequate diets are readily available commercially. However, the reviewer must be aware that many contaminants in the diet may have pronounced effects during gestation and lactation of rodents. Low levels of heavy metals, insecticides, mycotoxins, or synthetic estrogens may be present in the diet. These unintentional contaminants may occur naturally in plant materials or remain as residues from agricultural pesticide uses. Nitrosamine may be found in diets using fish meal as a major source of proteins, and aflatoxins may be detected in corn, wheat, and other cereals during storage. Many of these contaminants have resulted in serious effects on reproduction. They may act as teratogens (lead, mercury, cadmium, aflatoxins) or may prevent the implantation process as well as delay fetal growth (lead, cadmium; Degraeve, 1981). Furthermore, the reviewer should be aware that batch-to-batch differences exist in commercial food. Commercially available diets are formulated from natural products and thus are subjected to changes in nutrient composition as well as in contaminants. It should be possible to verify that the lot or batch of commercial chow used in the preparation of the diet throughout the entire investigation remained the same.

Nutritional requirements vary with the physiologic condition of the animals. Therefore, test animals usually require a higher intake of proteins, minerals, and vitamins during gestation and lactation. For example, the minimum concentration of protein content needed for maintenance of adult rats is 4.5% of the diet but increases to 12.0% during gestation and lactation (National Research Council, 1978). Vitamin deficiencies (E, riboflavin, thiamin) may be associated with infertility. Nutritional deficiencies are generally not a problem with contemporary studies conducted in major testing laboratories.

Regardless of the cause, the presence of a nutritional deficiency may result in many adverse reproductive outcomes such as (1) irregular cyclicity, (2) delayed puberty, (3) longer time

intervals necessary to become pregnant, and (4) loss of total reproductive lifespan (Ball et al., 1947; Berg, 1965; Frisch, 1978; Merry and Molehan, 1979; Nelson and Felicio, 1984). These nutrition-related alterations in reproductive performance may be reversible once the causative factor has been corrected.

d. Drinking Water

Contaminants that are found in the drinking water may be classified as suspended solids, organic solutes, or inorganic solutes (Shapiro, 1980). The suspended solids are mostly harmless; however, the organic solutes such as cyclic aromatics and halogenated hydrocarbons may exert an effect on the physiological response of the test animals. Organic contaminants which are frequently found in drinking water include halomethanes. These compounds derive from the interaction of a halogen (chlorine, bromide) and methane (from organic materials). The toxicological effects of many of these compounds, such as chloroform, have been well documented (IARC, 1990). Others have not been well investigated. Nitrate is another frequent contaminant of drinking water. Excessive levels of nitrates in water have been associated with methemoglobinemia.

e. Animal Housing

In a reproduction study, females are usually housed individually in solid-bottom cages except during mating. The type of bedding material should be reported since it may influence the biologic response of the animals. Wood shavings or chips are commonly used; however, hardwoods are preferred to softwoods since aromatic softwoods are well known hepatic microsomal enzyme inducers (Baker et al., 1979). Nesting material (cotton, shredded paper) is usually not necessary for rats.

f. Other Factors

The use of pesticides around the testing area is not recommended. Although sanitation of the testing area is crucial, pesticidal use may confound interpretation of the study results. Pesticides, air-deodorizing agents, and solvents may stimulate or inhibit the microsomal enzymes depending on the chemical used. Room deodorizers which consist of volatile hydrocarbons and essential oils may stimulate or inhibit the enzymes. Cleaning agents, solvents, and surfactants may have a similar effect. One disinfectant which is commonly used in the laboratory is ammonia. The reviewer should be aware that ammonia is an inhibitor of hepatic microsomal enzymes (Vessel et al., 1976). It is therefore suggested that non-chemical means of sanitation should be used. However, if the use of a solvent or insecticide is unavoidable, it should be clearly stated in the final report.

Antibiotics are sometimes needed to control infectious diseases during an on-going investigation. Anti-microbial agents may also have an impact on the physiologic response of animals. The use of any chemical in a reproduction study must be documented by the investigators and its potential impact on the study results should be carefully considered by the reviewer.

4. ROUTE OF ADMINISTRATION

In the testing of pesticides, the route of administration is generally through incorporation into the diet. Under special circumstances, dependent primarily upon the stability of the test material and its physical and chemical properties, it may be given by gavage, inhalation, dermally, or in drinking water.

Dietary exposure is the easiest route of administration since the animals do not have to be handled daily, and the amount of food consumed is directly proportional to the size and metabolism of individual animals. Dietary administration, however, is not appropriate for compounds which cannot be homogeneously mixed into

the diet or which degrade rapidly at room temperature. Therefore, analytical determination of the test substance in the diet must be performed periodically and appended to the final report. The homogeneity of the diet mixes, stability of the test material in the diet under storage and/or animal room conditions, and the frequency of diet preparation should be considered by the reviewer.

Gavage may be the recommended route of exposure if the oral route is indicated but administration through feed or water are neither practical nor appropriate. The nature of the vehicle and additives (suspension and wetting agents) must be indicated. The vehicle and additives used should not interfere with absorption of the test substance or produce toxic effects (EPA, 1982). Stability data of the test compound in the vehicle should be known. For insoluble test substances that are suspended in a vehicle, the reviewer should ensure that efforts were made to maintain homogeneity of the suspension during dosing.

Gavage dosing is commonly accomplished at a constant volume of 10 ml/kg of body weight. The dosing volume for each animal must be adjusted to individual body weight. In rodents, after the growth period (approximately 90 days), weekly adjustments are sufficient. The time of dosing is also of importance in a gavage study. Rodents are nocturnal animals and as such have their food consumption peak prior to day light. Administration of a test compound in the morning hours may result in a decrease in test material absorption due to the presence of food in the stomach. It is suggested that dosing by gavage, is preferably performed after mid-morning (Stevens and Gallo, 1982). Handling of animals during the dosing procedure constitutes an additional stress factor to pregnant animals and may lead to resorptions and/or abortions. Needless to say, the control group in a gavage study should be a vehicle control group receiving the same treatment as the groups administered test compound.

5. DOSE LEVELS AND DOSE SELECTION

At least three dose levels and a concurrent control should be used. Subdivision F of the FIFRA Guidelines (EPA, 1982) clearly indicates that:

a) The lowest dose level should not produce any evidence of toxicity.

b) The intermediate dose level should produce minimal observable toxic effects.

c) The highest dose level should produce some indication of maternal or adult toxicity.

Additional information on metabolism, pharmacokinetics, bioavailability, and/or bioaccumulation of the test substance should be available to demonstrate adequacy of the dosing regimen. It is recommended that steady state be reached prior to initiation of the mating period.

Among the objectives of a multigeneration reproduction study are to demonstrate a No-Observed-Effect Level (NOEL) and a Lowest-Observed-Effect Level (LOEL) for the parameters which have been investigated. The induction of minimal toxicity by the intermediate dose level is desirable but should not be viewed as essential. A LOEL is necessary to indicate that a sufficiently high dose level has been used to elucidate the potential of a chemical to cause reproductive effects. However, in the case that toxicity is not be demonstrated with the dose levels selected, the study may still be considered as acceptable if the highest dose level is at the limit dose (1000 mg/kg/day). Preferably, all dose levels used in a reproduction study should be reported on a mg/kg/day basis. However, in those cases where dose levels are provided on a ppm basis, the reviewer should convert the dietary concentrations to mg/kg of body using actual food consumption measurements or the nominal conversion factors reported in the Lehman tables (Table 2).

**[The Lehman table needs to be inserted here.]**

Table 2.

| | | | |
|---|---|---|---|
| | | | |
| | | | |

6. EXPOSURE PERIOD

The test compound is ideally given to the test animals on a seven-days per week basis. It is suggested that:

a) Parental (P) generation males and females should be exposed to the compound at a minimum of 6, and exposure should continue for at least 8 weeks prior to mating.

b) Female parental animals should be exposed to the test compound during gestation, lactation, the time interval during one or two successive matings (Fa and Fb), and until final sacrifice.

c) Offspring should be exposed without interruption from in utero, through lactation, weaning, and the growth period, and until sacrifice at weaning. Those selected to be parental animals of the next generation should be exposed through mating, the reproduction period, and until final sacrifice. The exposure period from weaning to mating should be at least 8 weeks.

Parental animals and their offspring should be continuously exposed to the test compound. Parental animals, especially males, in some studies which include two litters per generation, may be placed on control diet during the resting period (the interval period between Fa and Fb). This should be considered a study deficiency since all phases of spermatogenesis will not have been exposed to the test compound in the second litter.

7. MATING PROCEDURE

Paired mating (one male to one female) is preferred to colony mating in reproduction studies. Sibling matings must be avoided. The age of the animals at mating should be carefully checked to ensure that the animals have reached sexual maturity.

Mating is usually confirmed by the presence of a copulatory plug and/or the presence of spermatozoa in the vaginal smear. In the rat, vaginal smear examination for the detection of sperm is more reliable than the presence of a copulatory plug. The 1982 FIFRA guidelines indicate that unmated pairs may be remated with other proven sires or dams of the same group. However, this procedure, while possibly ensuring the production of a larger number of litters, may not contribute meaningful data for the evaluation of subfertility issues.

The individual animal data should allow the identification of the sire assigned to each dam during the first (or second) mating trials. The day of confirmed mating and delivery for each dam should be reported.

Both the OPP (1982) and OECD (1986) guidelines indicate that males and females should be cohabited until pregnancy occurs or until 3 weeks have elapsed. However, successful mating occurs within 4 days in approximately 90% of all pairings. If pregnancy does not occur in the allotted, possible causes of infertility in the pair should be considered. Information contributing to this analysis might include results of additional matings, female cyclicity data, sperm evaluations, or histopathological examinations of reproductive organs.

## 8. DATA REPORTING

Reporting requirements are listed in the Subdivision F Guidelines (EPA, 1982) and will not be discussed in detail in this evaluation procedure. In general, the final report should contain

tabulated data relative to parental body weight, parental organ weight, food consumption, parental mortality, reproductive indices, pup survival data and indices, pup body weight, male/female sex ratio, parental and pup necropsy data, and parental and pup histopathologic findings (Tardiff et al., 1977; Dixon, 1980). The reviewer should be able to identify the dam and sire from the individual animal data and to associate all reported findings with individual litter data. All reported mean data should be carefully compared to submitted individual litter data for consistency across all generations and dose levels.

## 9. STATISTICAL ANALYSIS

The statistical methods used must be described, referenced, and identified, since interpretation of reproductive results as well as of any toxicology data should rest on a sound statistical basis. The analysis of data arising in the reproduction study is complicated by the interdependence of various reproductive parameters and what has sometimes been referred to as "the litter effect". The latter is the result of the lack of independence of various observations observed in the same litter. The similarity of findings which is often observed among litter mates may be due to a variety of causes including genetic similarities, a common maternal environment, and differences in handling between dams. Techniques which have been developed to minimize "the litter effect" are discussed in Khera et al., (1989). Unless effects are clearly related to the male, the female should be used as unit for statistical purposes.

Statistical analyses performed in the Food and Drug Administration (Collins, 1978) use the two-tailed t-test for litter size, mean liveborn per pregnant animal, and mean pup survivors postnatally. Fertility indices are analyzed by the two tailed Chi squared test. Viability, weaning, and survival indices may be transformed by using the Freeman-Tukey arc sine transformation for

binomial proportions (Mosteller and Youtz, 1961) and are generally analyzed using the Dunnet's t-test. Data are sometimes reported on the basis of the proportion of the litters which are effected. The use of a nonparametric technique such as the Wilcoxon ranked sum test may provide a sufficiently powerful technique in the analysis of such data (Haseman and Soares, 1976). Indications of systemic toxicity, which are not complicated by litter effects, are analyzed using techniques such as Analysis of Variance and t-tests which are routinely used in the assessment of subchronic and chronic toxicity.

Different sets of statistical tests for reproduction studies may be used by other investigators and discussion of the appropriateness of each test is out of the scope of this evaluation procedure. When in doubt, it is suggested that the reviewer consult with HED statisticians.

10. FINAL REPORT

Submitted study reports should be signed and dated by the investigator(s); a signed quality assurance statement should be appended. This, of course, is not necessary for reports published in the open literature. If a study is not signed by the investigator(s) or if the histopathologic findings are not confirmed by a pathologist, it is assumed that the report is subject to change and does not represent the final position of the investigator(s). The reviewer should note that the report is considered as a draft and does not yet fully meet regulatory requirements, i.e., is classified no higher than Core Supplementary Data. The final report should be carefully compared to the draft report when the final report is issued.

## IV. STUDY INTERPRETATION

Most endpoints in the evaluation of a multi-generation

reproduction study can be grouped as either effects observed in the offspring or in the parents. However, this division (and the separation of endpoints into reproductive and systemic toxicity) is somewhat artificial because of the interrelationship of many of the aspects of toxicity which are observed in the reproduction study. Effects on fertility, for example, may have effects on litter size which may in turn influence pup weight and development and subsequent viability. The reviewer should not attempt to categorize the forms of toxicity which are observed except under certain circumstances (see Section I. C.).

## A. ENDPOINTS OF PARENTAL TOXICITY

In the evaluation of a multigeneration reproduction study, it is important to assess whether an adequately high dose level has been used. The highest dose selected should produce some indication of maternal or adult toxicity, and all systemic endpoints should be considered. These include significant changes in absolute body weights, weight gains, absolute and relative organ weights, feed and water consumption, clinical pathology, gross necropsy, histopathology, and cholinesterase activity data. No effects of toxicological significance should be observed on parental animals in the low dose group. The NOEL for systemic toxicity should be compared to NOELs in the chronic rat study to determine whether a greater sensitivity of the pregnant or lactating female is observed.

### 1. Clinical observations

Clinical observation data should include examination relative to the fur texture (matted, piloerection), skin (alopecia), eyes (mydriasis, miosis, nystagmus), mucous membranes (cyanosis), orifices (nasal discharge, vaginal bleeding), respiratory system (hyperpnea, dyspnea), autonomic and motor system (paralysis, paresis, fasciculation), behavioral changes (pica), and death. A

careful evaluation of the reported clinical signs should alert the reviewer to treatment-related effects and allow a more accurate determination of the NOEL and LOEL.

Parental death or body weight loss are obvious end-points of systemic toxicity and may result from many factors. Environmental factors, as discussed earlier, are known to influence the welfare of the test animals. Technical factors such as intubation error, e.g. perforation of the esophagus/stomach or intertracheal administration on a gavage study, and mishandling of the animals may alter the outcome of the test results and lead to maternal death and/or unnecessary additional stress superimposing pregnancy.

A necropsy should be conducted on all animals found dead during the course of the investigation. It is important to determine the cause of death if possible (accidental or compound-induced death) to better understand the clinical toxicity induced in the dams by the test compound. The pregnancy status and the time of death (pre-mating, mating, gestation, lactation, or rest period) should also be ascertained. Such information may be used to determine whether a selective toxicity is indicated during one or more of these periods. Maternal deaths, particularly at the high dose level, may influence the calculation of reproductive parameters and the interpretation of reproduction indices.

2. Parental Body weights

Parental absolute body weights and body weight changes are recognized as sensitive indicators of systemic toxicity for most species. Unfortunately, they are nonspecific and also may result from reduced palatability, rather than actual systemic toxicity. Decreased food consumption in the absence of other indications of toxicity suggests reduced palatability, and paired feeding studies may be necessary to identify the NOEL for systemic toxicity. The body weight data should be reported on a weekly basis (at a

minimum) for the growth period prior to mating as well as during gestation and lactation. Non-pregnant females should be excluded from most gestation and lactation calculations.

Body weight data are important for the determination of the NOEL. However, to be of greatest utility, all groups must have comparable initial body weights at the initiation of treatment (P generation). Decreased parental body weight gain is an effect which is usually seen in each generation, and it is recommended that trends in body weight be examined over the course of the entire study. However, it must be borne in mind that the P generation has received a shorter exposure to the test substance than the F1 generation and that their exposure did not encompass the period of perinatal development.

It has long been recognized that severe body weight loss can affect cyclicity in humans as well as in other mammalian species (Frisch, 1978; Merry and Holehan, 1979; Nelson and Felicio, 1984). However, a modest reduction in body weight gain as a result of decreased appetite is not expected to have any effect on reproductive parameters (Zenick and Clegg, 1989). A reduction in body weight may be due to a direct effect of the test material on the organism or may result from a decrease in food intake unrelated to the inherent toxicity of test material (decreased palatability). Therefore, body weight data should be assessed along with food consumption data for the calculation of food efficiency.

3. Food Consumption Data

If the test material is administered in the diet, the amount of food consumed is important in determining the exact amount of the test material received by the animal. In a multigeneration reproduction study, food consumption is measured on at least a weekly basis at different periods across all generations. The food consumption data is usually available for both parental males and

females during the premating period and for parental females during the gestation and lactation periods.

Although a reduction in food intake may be used to determine the NOEL, it is generally not observed in the absence of other indications of toxicity in the study. Food consumption varies from weaning to maturity, with younger animals consuming more food (on a kg/body weight basis). Unusually high food consumption data, which are actually due to spillage, are sometimes reported in toxicity studies. Very high consumption estimates in comparison to the Lehman tables (Table 2) suggest that spillage may have been a problem. The food consumption data during the lactation period are generally of questionable significance because of significant amounts of spillage due to the pups entering in the food container and also since the dam and her unweaned litter eat from the same container late in lactation (Collins, 1978).

Food consumption may be expressed as g/animal/day or g/kg body weight/day. The latter is preferred since the former excludes differences in body weights which may occur among the groups. As indicated earlier, body weight and food consumption data are best evaluated together since these parameters are interdependent. Evaluation of the body weight data along with the food consumption data may provide the reviewer with information relative to either a change in appetite (palatability) or a change in the food efficiency. The latter is a measure of the efficiency of the food utilization (food consumed per unit of weight gain). If the food efficiency index is similar between the treated and control groups, then anorexia may not be the main factor in depressing the body weight. It should be also noted that a decrease in food consumption or body weight may be due to other factors unrelated to the test chemical. Diarrhea, disease and decreased water consumption may be cited as some of the possible causes.

4. Water Consumption Data

Water consumption data are rarely available, unless the test material is administered via drinking water, since most test facilities utilize an automated water supply device which does not record individual consumption. In the event that water consumption data are recorded, an increase may suggest the possibility of renal toxicity. However, this finding should be corroborated by necropsy observations, histopathologic changes, or kidney weight changes to provide a conclusive endpoint of systemic toxicity. In rodents, a decrease in water consumption may lead to dehydration, decreased food intake, and body weight reduction, which eventually will lead to a sequelae of adverse reproductive outcomes.

5. Necropsy observations

The 1982 FIFRA Guidelines state that "a complete gross examination should be done on all animals, including those which died during the experiment or were killed in moribund conditions" and "special attention should be directed to the organs of the reproductive system". Uterine implantation sites can be counted to provide the means of estimating postimplantation loss via comparison to pup count at birth.

Necropsy data should be tabulated per group and generation. From these data it is essential to attempt to ascertain whether technical errors (e.g gavaging errors), diseases, or the test material toxicity itself, are responsible for the observed mortalities. For example, reddening of the trachea, congested lungs, and fluid accumulation in the lungs are highly suggestive of gavage errors and/or diseases. Further, not only the cause of death (accidental or compound-related) must be determined from the necropsy data but the pregnancy status of these animals must also be ascertained.

6. Organ Weight Data

Organ weights for neither the reproductive nor nonreproductive organs are required by the EPA or OECD test guidelines. As a result, they are often not available from reproduction studies for evaluation of either systemic or reproductive effects. However, organ weights are a useful component of the macroscopic examination, since they provide the first signs of dystrophic or dysplastic changes. Organ weights have been said to be of questionable significance unless the test material has a clearly specific effect(s) on a target organ. However, specific effects on target organs are not always known at the time of sacrifice, and measurements of organ weights are easily incorporated into study protocols. In the absence of organ weight measurements from reproduction studies, these data from other relevant and available toxicology studies may provide valuable information.

Organ weights should be expressed on both an absolute and relative basis. The relative organ weight takes into account the difference in body size (terminal body weight) since organ size increases with body size. However, the increase in organ weight is not directly proportional to body weight but instead is more closely related to surface area. The expression of the relative organ weight may not be biologically accurate in the presence of significant differences in terminal body weight among the groups. The interpretation of relative organ weights should be limited to groups of animals with comparable terminal body weights. In fact, the body weight reduction observed in the treated groups is frequently due to a reduction in fat deposition and not necessarily due to a depression in the development of lean body mass. Stevens and Gallo (1982) suggested that when significant treatment-related differences in a study are detected in many organs relative to body weight, organ/brain weight ratios should be analyzed, since subsequent to development the weight of the brain remains quite stable in adult animals.

Data on non-reproductive organ weight (e.g, liver, kidney,

adrenal glands, brain, spleen, and other known target organs) not only provides the reviewer with information relative to the target organ toxicity of the agent being tested but also is useful for the determination of the adequacy of dosing. Such information should be used in conjunction with organ weight data from subchronic and chronic studies in the same species to assess systemic toxicity.

7. Histopathological Findings

Histopathological information will be available for all high dose and control parental animals (P) and F1 animals selected for mating and in all organs showing macroscopic changes (EPA, 1982). If treatment-related histological findings are observed at any site, examination of those tissues from the mid- and low-dose levels is required. A list of tissues required for histopathologic examination is given in 83-4 of the FIFRA Guidelines (EPA, 1982).

Grading of the lesions is often necessary in the assessment of findings, particularly with commonly occurring histological changes. As for any other toxicological findings, a dose-response relationship in the frequency and/or severity of the findings assists in determining whether the effects are treatment-related. A compound-related effect may also be established if significant changes are observed only in the highest dosage level group. Microscopic lesions in the high dose group are usually specific to a small number of sites. Tissues from these sites in animals in lower dosage groups can then be examined in order to establish a dose-response relationship and a systemic NOEL.

Histopathological examination of the testes may include an evaluation of the spermatogenic process through identification of the 14 cell stage. This amount of detail is not necessary, however, if caudal epididymal sperm are evaluated for concentration, morphology, and motility. Serial histopathological examination of the ovaries to quantify oocytes may be performed but

generally provides little additional information without the concurrent presence of other hormonally-mediated effects in the females.

The liver and kidney, the two organs of primary importance for metabolism and excretion, usually show the highest incidence of pathological lesions.

8. Cholinesterase Determinations

Cholinesterase activity is often measured in the plasma, red blood cells, and brain of parental animals at sacrifice with known cholinesterase inhibitor pesticides. Cholinesterase measurements are rarely available for neonatal animals. If cholinesterase data are available and demonstrate pronounced biological significance in parental animals, they may be used as a basis for determining the NOEL for systemic toxicity and will be a factor in the decision as to whether the dose levels utilized were adequate for the investigation of reproductive toxicity. Measurements of cholinesterase in neonates or fetuses may indicate a preferential sensitivity of the young, as has been suggested by studies performed with aldicarb (Cambon et al, 1979). It has been suggested that the measurement of brain cholinesterase in pups be routinely conducted for pesticides with anticholinesterase activity (JMPR, 1990).

9. Endpoints of Parental Reproductive Toxicity

Reproductive indices are assessed for parental animals from mating to parturition and for pups from birth to weaning. Data generated from a multigeneration reproduction study provides information relative to the effects of an agent on germ cells, gametogenesis, libido, fertilization, implantation, embryonic growth and survival, fetal growth and survival, neonatal growth and survival, lactation, postnatal growth, and maturity.

a. Male Mating index

The male mating index is defined as:

$$\frac{\text{Number of males for which mating was confirmed}}{\text{Number of males used for mating}} \times 100$$

Although this index is often not reported, its calculation is feasible from the individual litter data submitted. It can be calculated for each generation and at each mating (Fa and Fb) of subsequent generations. It provides information relative to the number of sexually active males (ability to mate) in those studies where treated males are mated to untreated females. However, since both sexes are usually treated in multigeneration studies, this index is usually not a specific measure of male reproductive toxicity.

Mating is confirmed on the basis of the presence of vaginal plugs, of plugs in the pan beneath the animals, or of sperm in vaginal lavage. To truly ascertain the male mating index, pair mating is preferred. A decrease in the male mating index may be due to many factors, which include but are not restricted to, absence of libido, hormonal imbalance, or impotence of either sex. The etiology for a decrease in male mating index may be due to alterations in either the sensory, motor, hormonal or autonomic system. Regardless of the cause, evidence of a dose-response and persistent effect throughout subsequent generations is indicative of treatment-related effects in the male. A reduction in male mating index should be carefully evaluated along with possible histopathologic changes in the male reproductive organs.

Mating behavior parameters are useful because they can yield information about the integrative function of the neuroendocrine-gonadal axis. Evidence for an adverse mating behavioral effect in animals is considered suggestive of a potential for an adverse effect on human reproductive function.

b. Male Fertility Index

The male fertility index gives an indication of the outcome of mating and is calculated as follows:

$$\frac{\text{Number of males impregnating at least one female}}{\text{Number of males used for mating}} \times 100$$

In a rodent study, mating is normally determined by the presence of a plug (copulatory or vaginal) and/or presence of sperm in the vaginal smear. However, neither of these indicators necessarily indicates that pregnancy will ensue. In fact, a copulatory plug is only a product of secretions of the vesticular and coagulating glands of the male and does not necessarily indicate the deposition of sperm. The role of the accessory gland secretions is biologically unclear, since pregnancy may be induced in laboratory animals with sperm taken directly from the epididymis (Dixon and Hall, 1982). Likewise, the presence of sperm in the vaginal lavage does not imply that fertilization and/or implantation will occur. Further, the male rodent must provide an adequate number of intromissions and ejaculations for the female to respond with sufficient progesterone for the initiation of pregnancy (Aller et al., 1970; Chester and Zucker, 1970).

The male fertility index is also rarely reported by the study author(s) but can be calculated from the supporting data. The male fertility index provides information relative to the number of proven fertile males, but in studies in which both sexes are treated, it is not specific for male reproductive toxicity.

c. Copulatory Index

$$\frac{\text{Number of estrous cycles with copulation}}{\text{Number of estrous cycles required for pregnancy}} \times 100$$

Evidence of copulation is characterized by the presence of a copulatory plug and/or sperm in the vaginal smear. The number of estrous cycles or the length of time required for conception must be determined for each female from the supporting data. These data should contain information relative to male and female pairing, date of mating, number of copulations observed, and the number of estrous cycles required. Rats and mice usually allow mating to occur only at estrus (which lasts from 12-14 hours in 5 day cycles). Although rodents are spontaneous ovulators, ovulation does not accompany every estrous cycle, since estrus is dependent on gonadal hormones, whereas ovulation is responsive to gonadotropin (Jacoby and Fox, 1984). Normally, rodents require only one estrous cycle for mating to occur, and an increase in the number of estrous cycles that a female would require is suggestive of subfertility. Data on the estrous cycle may provide some indication of subtle changes in endocrine status, since the cyclic changes in vaginal cytology reflect the changes in endocrine milieu (Schwartz et al., 1977).

d. Female Fertility Index

This index gives a general measure of fertility of the strain and/or species, regardless of pregnancy outcomes, since females with evidence of mating (seminal plug or sperm) are presumed to become pregnant.

$$\frac{\text{Number of females conceiving}}{\text{Number of females cohabited with males}} \times 100$$

The FIFRA Guidelines (EPA, 1982) state that females unmated after 21 days (first mating trial) should be re-exposed to proven fertile males of the same treatment group (second mating trial). It should be noted that in many studies, re-mating of presumed infertile females may not be conducted, and the female fertility index may not be comparable.

The female fertility index is a measure of pregnancy, regardless of the outcome. The number of presumed pregnant animals (seminal plug or sperm) should be compared to the total number of dams that deliver, abort, or present evidence of fully resorbed litters. Therefore, a more accurate determination of this index as well as of the fecundity index (see section 1.e., below) requires an assessment of pregnancy by examination of the uterus of all animals which fail to deliver, for the presence of implantations and/or resorptions. Such an examination is usually carried out in a reproduction study at autopsy (after weaning of the pups); at that time the uteri of those dams may no longer display evidence of all implantation or resorption sites. Consequently, in many reproductive studies the reported fertility index may not truly reflect the number of pregnant animals, since the authors may report the fertility index as the percentage of females exposed to fertile males resulting in parturition. This calculation should be referred to as the birth index, which is defined as the percent of females exposed to fertile males which give birth to a litter of one or more pups. The birth index, in turn, should not be confused with the parturition index which is discussed later in this SEP.

Rodents not achieving pregnancy after the second mating trial should be considered to be infertile since they have had repeated opportunities to mate with fertile males. The use of proven males for a second mating is desirable because such use minimizes interpretation errors that can arise if inexperienced males are used.

e. Fecundity Index

This index reflects the percentage of matings resulting in confirmed pregnancies and is expressed as:

$$\frac{\text{Number of pregnancies}}{\text{Number of copulations}} \times 100$$

The fecundity index reflects the total number of dams that have achieved pregnancy, including those that deliver at term, abort, or fully resorb their litters. The main difference between this index and the female fertility index is that the fertility index measures ability of females to produce a litter when exposed to males whereas the fecundity index measures the ability to become pregnant after mating to a male.

As previously noted, one of the limitations of the multi-generation reproductive study is the inability to distinguish whether the reproductive adverse effects, if present, are primarily due to the male or female animals. A careful comparison of the male fertility index and female fecundity index and accompanying histopathology or other data, may provide the reviewer with some information.

As mentioned previously, copulation (presence of plug and/or sperm in vaginal lavage) does not necessarily ensure that fertilization and implantation will occur.

f. Gestation index

The gestation index is a measure of the efficiency of pregnancy resulting in at least one live offspring. The gestation index is defined as:

$$\frac{\text{Number of females with live born}}{\text{Number of females with confirmed pregnancy}} \times 100$$

This index is of limited sensitivity because litters with one or more live pups are treated as of equal biological significance. Therefore, although it is a measure of the number of litters with live offspring, it is an incomplete measure of fetal mortality unless the whole litter succumbs (Collins, 1978). To clarify the

effects of an agent on pup viability, it is suggested that the reviewer should concentrate on other ratios, such as the total number of pups per litter, the number of live pups per litter, and the number of liveborn per total number of pups born (Collins et al., 1975). The latter is a measure of the total number of offspring lost regardless of the total number of litters which may be available at term.

g. Reproductive organ weights

Although not mentioned in the 1982 Guidelines, data on reproductive organ weight should be collected for both male and female animals. Reproductive organs of interest are the ovary, and uterus for females and testis, epididymis, prostate, and seminal vesicle for males. In addition, adrenal gland weight may be a sensitive indicator of some hormonally-mediated effects.

Significant alterations in ovarian weight may suggest a female reproductive toxic effect. However, it should be noted that ovarian weight varies with the stages of the reproductive cycle and the number of corpora lutea present at sacrifice. Similarly, alterations in uterine weight should be evaluated carefully since the weight of this organ is under significant hormonal influence. Evaluation of uterine weight data in conjunction with data on the stage of estrus at sacrifice is more meaningful in approach.

Testicular weight data collected during chronic testing may provide an indicator of male reproductive toxicity, particularly those weights recorded at the time of the one year interim sacrifice. Testicular weights at final sacrifice in chronic studies are often complicated by a high background rate of atrophy. Although testicular weights do not exhibit a great variability between animals (Blazek, 1985), they are not a particularly sensitive indicator of testicular toxicity (Foote, 1986). Among the male reproductive organs, the testis shows the least weight

variation among normal members of a given species (Schwetz et al., 1980; Blazak et al., 1985). In the rat, testis grows at the same rate as the body. In addition, with senescence, there is a decrease in gonadal weight. The epididymis, prostate, and seminal vesicle weights may provide valuable information if care is taken in their removal and dissection. The accessory sex glands (seminal vesicles and prostate) can be weighed with or without fluids.

Pituitary and accessory sex organ weight should also receive attention as it may provide information regarding the target site or most sensitive endpoint for reproductive toxicity. Some physiologic functions of the pituitary gland are unrelated to reproduction. Hence, alterations in pituitary weights may be considered as adverse effects but do not necessarily reflect adverse reproductive effects (for either males or females). In general, changes in the pituitary are not considered to be more sensitive endpoints than those observed in other reproductive organs; therefore, lack of pituitary weight data does not usually compromise the study results.

h. Histopathology of reproductive organs

It should be noted that while histopathologic data may provide some information on reproductive function, they should not be used as conclusive evidence of an adverse reproductive outcome. Definitive conclusions can only be made with functional tests which analyze the response of the endocrine/reproductive system. Possible use of functional tests may be discussed in future Guidelines. In a reproduction study, morphological changes in both the endocrine and reproductive systems must be carefully monitored. Interpretation of histological changes at these sites is facilitated by recent publications such as that of Russell et al., (1991).

B. ENDPOINTS OF OFFSPRING (FILIAL) TOXICITY

1. Live Birth Index

This index is defined as:

$$\frac{\text{Number of pups born alive}}{\text{Number of pups born (total)}} \times 100$$

At birth, pups of all generations and groups should be examined for external anomalies as well as for viability. The number of viable, stillborn, and cannibalized members of each litter should be recorded (National Academy of Sciences, 1977). Distinction between stillborn (dead in utero) and pups which died shortly after birth (live born) is not always feasible since, from a practical point of view, the viability status of all offspring at birth could not always be performed immediately after parturition. However, this distinction could still be made by removing the lungs of dead pups and immersing them in water (the lungs of liveborn pups will float due to the presence of inhaled air). This distinction is of importance in case of whole litter death; assuming that all members of a litter are stillborn may modify the outcome of the gestation index, which is defined as the percentage of litters with one or more live pups. A decrease in live birth index, therefore, reflects compound-related effects manifested primarily during the advanced stages of pregnancy and resulting in stillborn.

Cannibalization is another problem which may arise and obscure the results obtained. Cannibalization is a behavioral change in the dams, which may be associated with general types of stress and is also a response by the dams to the delivery of malformed offspring. Inadequate food and/or water supply, elevated ambient temperatures, sudden changes in environmental conditions, lack of bedding, or poor animal handling techniques may be considered as part of the spectrum of stress-related changes (Harkness and Wagner, 1977; Cheeke and Patton, 1982). Cannibalization tends to be more prevalent among animals whose litters have developmental defects or are incompatible with life. Cannibalization may change

the "number of pups born" used in the calculation of the live birth index.

2. Viability Indices

The ability of the pups to survive is a primary focus in a reproduction study. Offspring viability can be impaired by:

i. Developmental effects of the young (abnormal and/or inadequate organ development) as a result of in utero exposure
ii. Varied effects of maternal toxicity:
    1. Maternal neglect (behavioral change)
    2. Inadequate milk production (endocrine change)
iii. Postnatal toxicity due to the presence of the agent in the milk during the lactation period

Offspring viability indices are usually measured on postnatal days 4, 7, 14, and 21, and each index represents a specific period of the animal life. The reviewer should note that the term "viability index" can be employed under two definitions. Fitzhugh (1968) refers to it as the percentage of all young born that are able to survive 4 days, but other investigators use this term to indicate the survival ability of the pups to other time points in postnatal life. Therefore, the meaning of "viability index" should be ascertained in connection with each report in which it is used.

For consistency, it is suggested that the viability index be used as a measure of 4-day survival:

$$\frac{\text{Number of pups alive on lactation day 4}}{\text{Number of pups born alive}} \times 100$$

Offspring deaths occurring at any time prior to day 4 may contribute to a decrease in this index. A careful examination of the individual litter data may provide the reviewer with some

indication of the etiology of the death. Early deaths in the offspring (i.e. post-natal day 1) are suggestive of functional defects (pulmonary, cardiovascular, or renal defect) rather than nutritional deficiency. Dead pups should be preserved and studied for possible defects and cause of death.

Data from some laboratories may indicate the presence of absence of a "milk spot" at the examination of moribund or dead pups. This is generally defined as a milk-filled stomach, which is externally visible through the thin, translucent skin of the pups. In addition, necropsy data may remark upon the presence of milk in the stomach. This information can be useful in distinguishing between deaths resulting from lack of maternal care or pup nursing ability and deaths resulting from toxicity to test substance in the milk or other causes.

On postnatal day 4, the litter size is sometimes standardized. It should be noted that although the 1982 FIFRA Guidelines (US EPA) indicate that all litters should be adjusted to 8, many other investigators prefer standardization to 10 to correspond with the number of mammary glands in rodents (3 pairs in the cervicothoracic region and 2 pairs in the inguinoabdominal region). However, it should be noted that lack of standardization does not necessarily imply that the study should be classified as Core Supplementary Data. In fact, many investigators believe that standardization is unnecessary and that it reduces study sensitivity (see section IV.

3. Lactation Index:

The lactation index is the viability index at postnatal day 21:

$$\frac{\text{Number of pups alive on day 21}}{\text{Number of pups alive on day 4}} \times 100$$

If litter standardization is performed then the denominator should read "mean number of pups kept after standardization on day 4". A decrease in the lactation index likely results from either in utero induced developmental effects, nutritional deficiency (endocrine change in the mothers), toxicity of the chemical (excreted in the milk), or maternal neglect. Regardless of the exact cause of death, an impairment of either the viability index or lactation index is clearly considered to be a reproductive effect.

4. Weaning Index (or 21-day survival index):

It should also be called to the reviewer's attention that in many non-standardized studies the weaning index is reported instead of the lactation index. The weaning index is a measure of:

$$\frac{\text{Number of pups alive day 21}}{\text{Number of pups born alive}} \times 100$$

The weaning index gives an overall offspring survival data from birth to weaning. To determine the overall offspring mortality in each litter, the reviewer should calculate the pre-weaning index which is defined as:

$$\frac{\text{Number of viable pups at birth} - \text{\# of viable pups at day 21}}{\text{Number of viable pups at birth}} \times 100$$

5. Litter Size

Mean litter size is another reproductive parameter that should always be considered in the evaluation of data. The mean litter size for each treatment group should be defined as:

$$\frac{\text{Total number of pups delivered}}{\text{Number of dams that delivered}}$$

In the calculation of the mean litter size, pregnant females that die or abort are excluded. It is more appropriate to use the total number of pups born (stillborn plus live pups) in the calculation of this mean instead of only the total number of live pups as reported by many investigators.

It represents the mean of live pups per litter based on the total number of females presumed pregnant. Determination of the pregnancy status of all dams is, therefore, crucial in calculating this value since the number of females presumed pregnant should include those which have had litters fully resorbed. A decrease in the live litter size may be indicative of a reproductive effect, resulting from a decrease in numbers of oocytes ovulated, an increase in the number of pre- or postimplantation losses, or an increase in the incidence of pregnant females which fail to deliver or have viable offspring. An increase in pre- and postimplantation loss may be the result of a genotoxic effect of the chemical. However, it is usually impossible to distinguish between failure of fertilization and death of the pre-embryo.

A decrease in litter size is indicative of possible adverse effects on either parental animal. Possible adverse effects in the males are determined by gamete quality and/or quantity. In females, adverse effects may have occurred during oogenesis, ovulation, fertilization, transport or implantation, and development of supportive organ systems (e.g. placentation).

A paternally-mediated effect on litter size can usually not be ruled out with the information which is routinely available in the reproduction study. If available, information on the number of oocytes ovulated (recently formed corpora lutea) and implantations is necessary in evaluating the extent of pre- and early postimplantation losses. These data may provide additional information that can be used in the overall evaluation of female reproductive effects.

6. Fertilization Efficiency (Pre-Implantation Loss), Implantation Efficiency (Post-Implantation Loss), and Fetal Gestation Viability Index

From the necropsy data of parental animals at weaning of the Fb generations, at least three indices can be calculated and are useful to study the effects of an agent on litter size. These are the fertilization efficiency, implantation efficiency, and fetal gestation viability indices.

The fertilization efficiency for each dam is defined as:

$$\frac{\text{Total number of implantations}}{\text{Total number of corpora lutea}} \times 100$$

and the implantation efficiency for each dam is a measure of:

$$\frac{\text{Total number of pups born (stillborn and live)}}{\text{Total number of implantations}} \times 100$$

Whereas the fetal gestation viability index is calculated as:

$$\frac{\text{Total number of live born pups}}{\text{Total number of implantations}} \times 100$$

Decreases in the fertilization efficiency and implantation efficiency indices represent, respectively, preimplantation loss and postimplantation loss. An evaluation of the above indices provides information concerning effects on fertilization, implantation and early and late deaths of zygotes.

7. Pup Body Weights

In addition to the above offspring indices, a very important measurement of reproductive toxicity is the weight of the surviving pups. The 1982 FIFRA Guidelines suggest that live pups should be counted and litters weighed, by weighing each individual pup

(optional) at birth, or soon thereafter, and on days 4, 7 (optional), 14, and 21 after birth.

Pup body weight data should be evaluated concurrently with pup survivability data. As discussed earlier, an increase in offspring mortality after postnatal day 4 may result from either nutritional deficiency (hormonal imbalance in the mothers), maternal neglect (behavioral change), or directly from the toxicity of the chemical tested (excreted through the milk). Therefore, an increase in offspring mortality without impairment of pup body weights may, at least, rule out the possibility of nutritional deficiency, whereas concurrent increases in offspring mortality and decreases in pup body weights may result from all these factors. Regardless of the exact etiology of increased mortality and decreased pup body weights, these findings should generally be considered as toxic effects. Like other toxicological parameters, if an inconsistent but statistically significant decrease in pup body weight is found in the treated groups, the data should be compared not only with the concurrent control but also with control data from other generations and recent historical control data.

The weight of the pups at weaning (lactation day 21) is another important parameter that should be considered in the evaluation process. A difference in neonatal birth weight does not necessarily imply that a difference in weaning weight will ensue. The weaning weight may be biologically and statistically similar to controls (reversible effect) or remain altered (irreversible effect). However, attainment of expected weight at weaning does not demonstrate that untested functional effects have neither occurred nor persisted. It is also important to keep in mind that litter size has an important influence on pup weight. Mean pup weight shows a slight but consistent decrease with increasing litter size when litters are six pups or greater (Khera, K. et al., 1989). If litter size is greater in treated groups than in the control group, whether due to chance or as a compound-related

effect, decreased pup weights may be expected (HED Ad hoc Committee for Atrazine Reproductive Issue, 1992) for a discussion of the setting of a NOEL in a reproduction study with varying mean litter sizes

8. Crown-rump lengths

Crown-rump length measurements are not required by the 1982 FIFRA Guidelines. However, if available, these data are useful for evaluating offspring growth development in conjunction with the body weight data. Crown-rump length measurements are usually well-correlated with body weight.

9. Developmental milestones

The evaluation of developmental parameters in the study offspring, although not required by the 1982 FIFRA guidelines, can provide useful information regarding sexual maturation as well as indications of hormonally-mediated effects on the offspring. Examination of pups immediately after birth for alterations in anogenital distance may help identify early subtle signs of reproductive toxicity such as that seen with procymidone. At day 14 of lactation, examination of the pups for normal development of nipple structure would further aid in identification of feminized males. The onset of sexual maturity, generally a body weight-dependant milestone, can be evaluated by examination of females for the opening of the vaginal orifice, and by determination of preputial separation in the males. Delays in these normal processes, whether or not linked to body weight, may indicate a toxic response to treatment.

10. Necropsy and histopathology of offspring

Subdivision F of the 1982 FIFRA Guidelines (USEPA, 1982) states that "a complete gross examination should be done on all

animals, including those which died during the experiment or were killed in moribund conditions". Although this wording implies that all offspring should be necropsied, the National Research Council (1977) suggested that only 10 male and 10 female pups, randomly selected from the F2b litters of each test group and the control group, should be sacrificed at weaning and subjected to a complete gross examination. An alternative approach includes the random selection and necropsy of one pup per sex per litter of the F1 weanlings not destined to become parents of the second generation. In this study design, F2 pups might not be examined by gross examination unless changes were observed in the reproductive organs of the necropsied F1 weanlings. In the absence of guideline recommendation that clarify this issue, all approaches are considered acceptable.

Neither organ weight nor histopathology data of the pups are mentioned in the FIFRA Guidelines (USEPA, 1982). One suggested approach to postmortem evaluation of weanlings includes weighing the ovaries, testes, brain, liver, kidneys, and known target organs from those pups selected for gross necropsy. The ovaries, testes, target organs, and grossly abnormal tissue would be preserved for histopathological examination. Organs demonstrating treatment-related changes in the weanlings would be examined histopathologically for the control and high-dose groups. In this approach, the F2 weanlings would not be routinely examined because they would not be expected to provide a source of new or different information. A lack of offspring histopathologic examination should not be used as a basis to classify a study as Core Supplementary Data.

11. Use of Historical Control Data

Reproductive performance varies widely within the same species due to genetic and environmental factors. Although concurrent control data are normally the most appropriate for comparison to

data from treated groups, on occasion, historical control data for the same strain/species provide the reviewer with valuable information regarding the background rate for various reproductive parameters as well as normal variations and trends. **In general, historical control data should not be used as a substitute for concurrent control results and need not be routinely requested. However, testing laboratories should be encouraged to routinely collect historical control data.**

The reasons for requesting historical control data for reproductive toxicity studies are similar to those for developmental toxicity studies:

a. Historical control data can provide a guide for determining the biological significance of statistically significant differences observed in a reproductive study. Such data may indicate whether a concurrent control group incidence (e.g., pregnancy rate, dams with 100% resorptions, etc.) are unusually low for the test species, thereby artificially enhancing the statistical significance of findings in treated groups, or whether nominally increased incidences for a treatment group are within the normal range seen for that strain/species. While the range of reported historical values is important, it is presumed that individual findings outside two standard deviations of the mean are outliers and therefore may not be acceptable for comparison against treatment groups in the interpretation of statistical versus biological significance.

b. Historical control data may also indicate trends in the overall vigor, fertility or litter size of a particular strain/species which relate to genetic drift and can help in the interpretation of apparent unusual findings.

Some specific examples of situations when historical control data would be necessary to clarify the findings in a reproduction study are presented by category:

a. Unusual concurrent control findings, leading to difficulty in interpretation of the validity of significant findings in treated groups

i. A clear dose response is not evident for observations such as decreases in fertility, increased incidence of abortions or premature births, smaller live litters, or number of dead pup/litter; however, the values for one or more treatment groups are elevated above the concurrent control such that statistical or apparent biological significance is achieved.

ii. Fertility appears to be unusually low in the concurrent control and/or treatment groups and is suspected of impacting the overall validity of the study.

c. Variation between generations

i. Differences in various reproductive parameters, e.g., litter sizes or mean pup weights, are noted between control generations (F1, F2).

ii. If developmental landmark timeframes are recorded (e.g., anogenital distance, vaginal opening, or preputial separation), they may vary between generations, and this variation may appear to be biologically or statistically significant.

c. Unusual findings

i. Developmental anomalies or variations are noted in a treatment group but not the concurrent control, and the findings are not dose-related. Background incidence data from either reproduction or developmental toxicity studies may assist in determining biological significance.

ii. An apparently high incidence an observation such as pup mortality is observed at all dose levels, including control, although no significance is demonstrated in treated groups.

Examples of the types of historical control data necessary to effectively evaluate a reproduction study are given in Table 3. These represent the most critical information for a complete Agency interpretation of reproductive toxicity studies but are not a comprehensive listing of reproductive indices or study parameters. In order to facilitate study evaluation, the reviewer may request that other historical control data, not included in the Table, be provided by the registrant.

For both developmental and reproductive toxicity studies, it is preferable that the initiation of treatment for the first generation of studies used in compiling specific historical control data sets be within a range of ± two years of the initiation of treatment for the study under review (Developmental Toxicity SEP, 1992). However, studies for which the initiation of treatment is within ± five years of the study under review are generally considered to be acceptable for inclusion into the historical control data set for reproduction studies. This differs from the criteria for developmental toxicity historical control data because the types of reproduction data requested (see Table 3) require longer times to generate. In addition, they are not as sensitive to subtle changes in examination procedures or diagnostic criteria which may affect the type of abnormalities recorded in

developmental toxicity studies. Furthermore, a wider range in allowable larger time frame allows for potential genetic drifts in the strain to be noted. Needless to say, due to the very real possibility of genetic drift, historical control data derived from studies conducted closer to the date of the study under review should be given more scientific weight than studies conducted at either end of the five-year time range limit.

Studies which are not representative or typical, either in study design and conduct, or in the condition (health) of the animals, should not be included in the historical data set. Individual study results as well as summary statistics, i.e., mean, median, standard deviation or error, and/or range values, must be included in the historical control data submitted to the Agency. Individual studies should be identified by dates of initiation and completion of the in-life phase, vehicles (if any) utilized, method of administration, changes in feed or animal suppliers, environmental conditions or other significant factors which change over time. Data sets should be separated and identified by generation.

Historical data sets derived from multigeneration reproduction and/or developmental toxicity studies conducted in multiple laboratories and published in the scientific literature (Clemens et al., 1992) may be considered by the reviewer in the interpretation of study results. These data may be particularly useful when the performing laboratory is unable to provide a historical control data set from their own facility, although such a situation is rare. Generally, however, these data must be regarded with a certain amount of caution, since there may be unknown variability in study conditions and conduct between various laboratories, as well as inapparent differences in technical procedures, interpretation of effects, and calculation of indices.

Table 3. Examples of Multigeneration Study Data Which Might be Included in a Historical Control Data Set[a]

| |
|---|
| **General Information** |
| Number of males and females mated |
| Number of females pregnant |
| Number of confirmed pregnancies/number of confirmed matings |
| Number of males and females surviving to scheduled termination |
| Number of females with abortions |
| Number of females with premature deliveries |
| Number pregnant up to parturition |
| **Adult Reproduction Indices[b]** |
| Copulatory Index |
| Mating Index |
| Fecundity Index |
| Male Fertility Index |
| Female Fertility Index |
| Parturition Index |
| Gestation Index |
| **Pup Reproductive Indices and Parameters[b]** |
| Live Birth Index |
| Sex Ratio |
| Live Litter Size Index |
| Sex Index |
| 4-Day Survival Index (Viability Index) |
| 21-Day Survival Index (Weaning Index[d 0-21]/Lactation Index 4-21]) |
| Preweaning Index |
| Mean pup weight at: |
| Day 0 (birth) |
| Day 4 (pre-standardization) |
| Day 4 (post-standardization) |
| Day 7 |
| Day 14 |
| Day 21 (weaning) |
| Incidences of developmental anomalies |

a Historical control data for each generation of a given study should be included.

b For definitions, see discussion on reproductive indices and parameters (section ). Consistency between studies in calculation of indices is required.

C. Other Considerations in Study Interpretation

As in any toxicity study, the design of a multigeneration reproduction study has certain limitations.

1. The inability to associate effects with gender

One of the most important limitations of the reproductive data is the inability to clearly identify male and female adverse reproductive effects. Both males and females are routinely treated in reproductive studies, and changes in reproductive indices reflect the contributions of both sexes. Although suggestions of gender-specific effects may arise from organ weight changes and histological examinations conducted in this study (or from evidence in other studies such as chronic feeding or developmental toxicity), conclusive evidence of gender-specific toxicity requires testing beyond that routinely required by the Office of Pesticide Programs. A male-mediated effect may be more easily shown if histopathology and sperm evaluations are requisites of a study design. In the presence of probable male-mediated effects, a request for a "crossover mating" study, in which treated males are mated to untreated females, is sometimes necessary. However, regulatory action by the EPA has increasingly become gender neutral (i.e. designed to protect both sexes), and determination of gender sensitivity may thus not be essential.

2. The lack of specificity and sensitivity

It should also be borne in mind that many reproductive indices lack sensitivity and specificity. For example, parturition is not always well monitored since rodents, in general, deliver nocturnally. The litter size is only an estimate, since a correct litter size should include cannabalized pups, which are not readily accounted for in a reproduction study. The effects on germ cells, gametogenesis, libido, implantation, and embryonic growth and survival are only indirectly measured in the reproduction study (Schwetz et al., 1980).

The identification of adverse effects in the reproduction study should not be limited to observations on infertility, adverse

pregnancy outcomes and adverse effects on offspring survival and growth but should be extended to (i) subtle alterations in structural or functional competence of the ovaries, (ii) feed back mechanisms, (iii) onset of puberty, (iv) vaginal cytology, (v) premature reproductive senescence, (vi) histological evaluations of accessory sex glands, (vii) histological evaluations of spermatogenesis, (viii) endocrine evaluations, (ix) biochemical markers, and (x) evaluations of sexual behavior.

a. Sperm evaluation for male-mediated effects

Sperm measurements may provide valuable information in cases where reproductive toxicity is male-mediated. Although sperm measurements are most readily obtained from species larger than rodents, data concerning sperm production and characteristics are increasing being gathered from subchronic, chronic and reproduction studies in rats, and it is important that this information be carefully considered in the evaluation of reproductive toxicity.

It is preferable to use caudal epididymal sperm (collected at necropsy) for evaluation rather than samples taken from electroejaculation. The latter technique may result in sperm measurements that are not reflective of the sperm that may be delivered through normal copulation. If a treatment-related effect is observed in the caudal epididymal sperm concentration, the spermatid count may be evaluated.

The production of sperm may have to be severely reduced (e.g. by 80 to 90% in some strains of rodents) before fertility is effected in the multigeneration reproduction study (Robaire et al., 1984). Human fertility, on the other hand, may be effected by a small decrement in sperm production at least partially because many men over the age of 30 have daily production rates of normal sperm which are inadequate or barely adequate to ensure fertility (Amman, 1981). Direct measurements of sperm production and function in rodents may therefore be highly sensitive and specific indicators

of toxicity. An additional advantage of performing sperm measurements is that this data can also be obtained from humans in situations where animal studies suggest a hazard, enhancing the ability to confirm the extrapolation of test results to humans. Two reviews of sperm evaluations have been published by the EPA Gene-Tox Program (Wyrobeck, 1983a; 1983b). Other reviews of the use of sperm measurements in rodents include Amann (1986) and Blazek et al. (1985). These references should be consulted for detailed guidance on the interpretation of sperm measurements and for a compilation of reference values. However, the reviewer should be aware of the most frequently collected sperm measurement information. This includes sperm count (as an indicator of sperm production), sperm morphology (which generally includes only the evaluation of head shape and is thought to be related to the ability of a sperm to reach and fertilize the oocyte) and sperm motility, expressed as percent progressively motile sperm (which is influenced by a large number of variables and is an indicator of both sperm maturation and ability to fertilize). The use of video technology in sperm assessment is preferable, since it provides a permanent record of the sample collected and allows more accurate measurement of the data. General species differences for sperm parameters are shown in Table 4.

Table 4. Species differences in spermatogenesis, daily sperm production, and epidiymal transit time[a]

| Parameter by species | Mouse | Hamster | Rat | Rabbit | Dog (beagle) | Monkey (Rhesus) | Man |
|---|---|---|---|---|---|---|---|
| Duration of spermatogenesis (days) | 34-35 | 35-36 | 48 | 48-51 | 62 | 70 | 72-74 |
| Duration of cycle of seminiferous epithelium (days) | 8.9 | 8.7 | 12.9 | 10.7 | 13.6 | 9.5 | 16.0 |
| Life span (days) | | | | | | | |
| B spermatogonia | 1.5 | 1.6 | 2.0 | 1.3 | 4.0 | 2.9 | 6.3 |
| Leptotene | 2.0 | 0.8 | 1.7 | 2.2 | 3.8 | 2.1 | 3.8 |
| Pachytene spermatocytes | 8.0 | 8.1 | 11.9 | 10.7 | 12.4 | 9.5 | 12.6 |
| Golgi spermatids | 1.7 | 2.3 | 2.9 | 2.1 | 6.9 | 1.8 | 7.9 |
| Cap spermatids | 3.6 | 3.5 | 5.0 | 5.2 | 3.0 | 3.7 | 1.6 |
| Testicular wt. (total) | 0.2 | 3.0 | 3.7 | 6.4 | 12.0 | 49.0 | 34.0 |
| Daily sperm production (millions) | | | | | | | |
| Per gm of testis | 28 | 24 | 24 | 25 | 20 | 23 | 4.4 |
| Per male | 5.6 | 72 | 89 | 160 | 240 | 1127 | 150 |
| Sperm reserves in caudae epididymides at sexual rest (millions) | 49 | 1020 | 440 | 1600 | 2100 | 5700 | 420 |
| Epididymal transit time at rest (days) | | 14.8 | 8.1 | 12.7 | 11.3 | 10.5 | 5.5-12 |

a Data derived primarily from Amann (1986) as cited in Biologic Markers in Reproductive Toxicology (NRC, 1989).

b. Cyclicity data in the evaluation of female-mediated effects

The evaluation of female cyclicity data can be helpful in identifying female-mediated effects on fertility. Samples of vaginal epithelium are collected daily by lavage and examined microscopically to determine the stage of estrus.

Although it is possible to collect cyclicity data daily throughout the study, evaluation of specific points in the reproductive life of the study animals will yield sufficient data for evaluation of effects. The onset of female cyclicity can be determined by daily smears of puberty-age females; this milestone in reproductive development can also be evaluated by examination ofor the age at which vaginal opening occurs. Evaluation of

vaginal smears during the mating period can sometimes offer explanations for a female-mediated fertility reduction in the mating pair. For females that show no evidence of copulation while paired with a male, the duration of the mating period can be defined or regulated by the number of full or partial estrous cycles that pass. Knowledge of the stage of estrus at the time of necropsy can aid in the interpretation of female necropsy and organ weight data, particularly for uterine observations, since the turbidity of the uterus is hormone-dependant. Taking this procedure one step further, some protocols may even require that all females be sacrificed at the same stage of estrus in order to facilitate interpretation of the data.

The reviewer should be aware that care must be taken in the laboratory to avoid stimulation of the cervix with the pipette during the vaginal lavage procedure, or pseudopregnancy (as evidenced by prolonged proestrus) may be inadvertantly induced, and fertility may be artificially diminished. This lack of good laboratory technique could potentially compromise the results of an otherwise adequate study.

3. Limitations of Test Sensitivity

Reproductive performance varies widely within the same species due to genetic and environmental factors. Although concurrent control data is normally the most appropriate for comparison, historical provides the reviewer with valuable information regarding the background rate for various reproductive parameters as well as normal variations and trends. Historical control data should, of course, be for the strain/species tested from the same testing facility and should have been generated within a recent time frame (no more than 5 years from the time of the study in question). Data should be presented on a study by study basis. When using historical control data, the reviewer should be aware that there will be some difference between animals of the same

strain from different suppliers and that the extent of inbreeding of commonly used laboratory rats has resulted in certain long-term trends. The Charles River Sprague-Dawley rat, for example, has been heavier, shorter-lived, and more fecund (with larger litters). A more complete discussion of the use of historical control data can be found in section IV.D of this document.

4. Limitations of Data Extrapolation

For the same chemical, there are often species differences in both pharmacokinetics and activity at the target site. Comparative pharmacokinetic data has rarely been available for pesticides but may explain much of the difference in the response of various species, genders and strains to test chemicals. Furthermore, with the advent of physiologically-based pharmacokinetic modeling, comparative pharmacokinetics may allow more accurate extrapolation between species. In the absence of good predictive data regarding concentrations of the proximate toxicant at the target site (testes, embryo, etc.) in man and in the test species, it is generally assumed that humans and test species are similar in the pharmacokinetic disposition of the chemical and in the response of the target tissue. Differences in species sensitivity at the target site has been documented less frequently but may not be an important factor for chemicals with steroidogenic properties.

a. Poorly Defined Indices

Poorly defined indices often tend to cloud or bias findings (USEPA, 1987). Summary data can be useful but should always be submitted in conjunction with individual untransformed data sets to allow evaluations of individual end points. In addition, indices have many definitions, and the reviewer must carefully examine these to assure there is meaningful and consistent description and that this is correctly presented in study reviews.

b. The Relevance of the Animal Model in the Detection of Effects on Fertility

The rat is not an ideal surrogate in studies which are designed to assess the potential for human reproductive effects. For example, the ejaculate in the rat generally contains a large excess of spermatozoa and a small reduction in sperm count would not likely be detected in mating trials. In contrast to rats, humans lack this large excess capacity. A reduction in sperm count which has no effect in the rat may result in functional sterility in the human male (Johnson, 1986). There is the opportunity for repeated copulations during each cohabitation in the multigeneration reproduction study and although the likelihood of fertilization may be reduced for each mating, repeated matings may minimize the sensitivity of study to detect adverse effects on fertilization. It has been suggested that limiting the matings to one per cohabitation or limiting the length of the mating period (to one week or, alternatively, for the duration of 3 full or partial estrous cycles) may increase the sensitivity of the multigeneration reproduction study.

c. Inability to Detect Effects Upon Functional Reserve Capacity in Neonates

Perinatal and postnatal functional maturation could be partially impaired but remain undetected during clinical observations or through examination of indices assessing viability. For example, exposure to a toxicant during fetal stages may result in diminished respiratory reserve capacity which would only be detected in the conduct of special testing. The animal may look and function normally, but, when challenged, may show a decreased response. Its lungs may not fully inflate, the septal walls may be inadequately attenuated, and oxygen consumption may be depressed (Johnson, E.M. (1986), The Scientific Basis for multigeneration Safety Evaluations, J.Am. Col.Tox., 5:197-201; Newman et al. 1984;

Newman et al. 1983). Adverse effects on reserve capacity after in utero exposure may also be observed on the renal or immunological systems (Kavlock ?, etc.). Effects such as these may not be discovered in routine multigeneration reproduction studies.

d. Recessive Gene Effects

Palmer (1981) cites the following (Table 5) as an example of an increased incidence of abnormal offspring in F1 and F2 generations which was not due to treatment but rather to the inheritance of a recessive gene as seen through the careful examination of derivation records through the three generations.

Table 5. Effects of a recessive gene in a rat multigeneration study (incidence (%) of litters containing offspring with locomotor incoordination)

| Treatment | P Generation | F1 Generation | F2 Generation |
|---|---|---|---|
| Control | 0.0 | 2.6 | 0.0 |
| Low | 0.0 | 0.0 | 8.2 |
| Intermediate | 13.0 | 2.7 | 6.7 |
| High | 0.0 | 6.8 | 7.1 |

The gene was only expressed in the intermediate dose level in the P generation, since it was only at that dose level that one or more matings of heterozygous parents occurred. Such matings occurred more frequently in the two subsequent generations due to chance, and the low background expression of this gene clearly does not reflect a compound-related effect.

## IV. Issues Concerning Study Design

### A. The One Generation vs. the Two Generation Study

The EPA held a workshop (1987) to discuss whether it was

necessary to continue to require a two-generation study in light of a review study by Christian (1986). One of the conclusions of the study by Christian was that, "In the absence of evidence of bioaccumulation, the studies reviewed provide limited evidence that a one-generation reproduction study is a sufficient test for determining the no-effect level for reproductive toxicity."

Table 6 summarizes three independent comparisons of effects observed in the first and second matings of the first and second generations. The table indicates that toxicity is most often seen at the time of the first mating (first litter) and has been used as justification for modifying the design of the multigeneration study to include only a single mating. It is also of interest that effects are often seen in the first mating of the parental generation although the treatment duration is less than in subsequent generations.

Table 6. Detection of first effects in selected reproduction studies

| Generation | Mating | No. of studies[a] | | |
|---|---|---|---|---|
| | | Clegg[b] | HRC | Christian |
| P | 1st | 32 | 23 | 20 |
| | 2nd | 5 | 0 | -- |
| F1 | 1st | 5 | 2 | 0 |
| | 2nd | 0 | 0 | -- |
| F2 | 1st | 2 | 0 | 0 |
| | 2nd | 0 | 0 | -- |
| Studies with no effects | | 27 | 9 | 35 |

a The number of studies in which an effect was first detected in a given generation or mating.
b Authors of comparison studies.

Although the investigation discussed above notes that chemicals which bioaccumulate may require a second generation, adequate pharmacokinetic data which would provide evidence of bioaccumulation are rarely available prior to the initiation of

multigeneration reproduction studies. It was generally agreed, however, that it would be beneficial if better information concerning pharmacokinetics were available prior to initiation of reproduction studies as it could, for example, enable one to modify a study design to ensure that plasma levels in animals have achieved a steady state condition prior to mating. Such an approach may be necessary with organochlorine compounds and other compounds with long half-lives for elimination. A workshop sponsored by OPP was held at the National Academy of Sciences (1992) to discuss the use of pharmacokinetic data and risk assessment. A tiered approach to the gathering of pharmacokinetic information was proposed in which the first tier of information would be developed early in the testing of a compound. The first tier would consist of limited investigation which would yield half-lives for elimination and other basic information. Additional information, such as pharmacokinetic data from pregnant animals, would be generated if triggered by concerns in the area of reproduction or developmental toxicity.

Bioaccumulation is not the only factor which may account for effects seen after the first generation. Effects in parental animals of the second generation may be due to the fact that the F1 animals are exposed <u>in utero</u>, via lactation, and directly from the time of weaning, whereas exposure of the P generation is not begun until the animals are 6 to 8 weeks of age. The ability of perinatal rodents to metabolize agents is limited due to a lower level of mixed function oxidase and conjugation capacity compared to adult animals. It is not until the animals are 30 to 60 days of age that metabolism and excretion capabilities approach that of an adult (USEPA 1987).

In summary, although most studies will demonstrate reproductive effects after the first mating, some chemicals will show effects at lower levels, or qualitatively different effects, after the second mating due to exposure during a longer portion of

the lifespan, and, for chemicals with long half lives for elimination, more time to achieve a steady state level. For these reasons, a second generation is required in the standard protocol for a reproduction study.

B. The Need for a Second Litter in each Generation

Bioaccumulation is also relevant to the need for a second litter in each generation. Chemicals with very long half lives for elimination (such as DDT) may not reach a steady state concentration at the time of mating for the first litter and the second litter may be more sensitive for the detection of reproductive effects for these chemicals. The duration of treatment of the parental animals is also longer at the time of the second mating, the animals are older and larger. All of these factors may lead to the second litter being more sensitive than the first litter. However, the need for a second litter in multigeneration studies should be considered on a case-by-case basis and should not be required without justification. A second litter was routinely included in studies conducted prior to the mid-1980s.

C. Length of the Premating Period

The length of the pre-mating treatment period has ranged in various protocols from 8 to 14 weeks. The 1982 OPP test guidelines suggest, but do not require, a premating period of 14 weeks rats. The OECD test guideline for a reproduction study and that of FDA recommend a minimum of 10 weeks. The premating period must be at least long enough to cover a sperm cycle plus epididymal transit time in males and a period of 5 estrous cycles for females. Because the female rat cycle is only 4-5 days, the limiting factor in the necessary dosing period prior to mating is ensuring test compound exposure throughout all periods of spermatogenesis. In some cases, pregnancy rate may be related to the length of the

pre-mating period with longer premating periods resulting in a lower and highly variable fertility (Palmer, 1981). A short premating may result in younger animals of low body weight which have low birth weight pups. Although the optimum period will vary slightly among strains of rat, the 10 week minimum period recommended in the OECD test guidelines should be considered as sufficient for commonly used strains of rats.

D. Standardization of Litters

The 1982 OPP test guidelines recommend standardization (culling) of litters to 4 males and 4 females on day 4 postpartum. As previously noted in Table 1, OECD test guidelines make the culling of litters optional, and in on-going efforts towards harmonization of OECD and USEPA guidelines, this approach has taken precedence. The FIFRA Accelerated Reregistration Phase 3 Technical Guidance (December 24, 1989) indicates that standardization of litters should not be considered as mandatory.

Some arguments for not standardizing litters have been advanced by Palmer (1986). Among the list of issues raised are the following:

1. Standardization disrupts the normal distribution of litter sizes.
2. The standardized litter size is below the natural mean, median and modal values which are normally observed.
3. Standardization results in the alteration of, on average, 77 percent of normal litters in young rats, and 91 percent of litters in older rats.
4. Standardization results in the elimination of 25 to 40 percent of normal offspring, most of which would survive.
5. Human bias or error is introduced in the selection

of retained animals.

6. The introduction of bias preferentially acts against controls.
7. Standardization raises mean body weight at weaning and thus reduces the likelihood of discovering a lactational effect on body weight.

A common belief about the use of standardization has been that it helps in obtaining more consistent mean pup weights. Although the mean weight may be more consistent throughout weaning after standardization, effects on mean pup weight and mean litter weight may not be as readily detected after standardization due to the loss of information which would have been obtained from the culled pups. Mean pup weight and mean litter size are the two factors which determine the mean litter weight. The mean litter weight is most analogous to neonatal weight in monotocous species such as the human.

It should be noted that many reproductive toxicologists prefer standardization of litters. Variability is decreased and consistency in litter size facilitates statistical analysis. Unusually large (or small) litters may have consequences on mean pup weight and survival and may complicate the identification of compound-related toxicity on these parameters. In addition, it has been stated that study sensitivity actually increases with standardization. This latter assertion has not been tested experimentally or statistically.

## V. Weight-of-Evidence Determinations and Risk Assessment

The basic information to be evaluated in a reproduction study has been reviewed in previous sections of this SEP. Section A of the risk assessment discussion outlines the information necessary for presentation to the Health Effects Division Reproductive Toxicity Risk Assessment Peer Review Committee (PRC). This

approach is similar to that utilized for developmental toxicity and is consistent with the Agency Reproductive Guidelines. Section B attempts to place in perspective certain frequently encountered toxicological and exposure issues which must be considered in evaluating the reproductive toxicity potential of pesticidal chemicals. Finally, in Section C current and future risk characterization approaches are discussed.

A. Outline for Peer Review Committee Presentations and DERs

The reviewer should refer to the Guidelines for Assessment of Reproductive Risk (USEPA, 1992?) when preparing a Data Evaluation Report (DER) or Peer Review Document for Reproductive Toxicity. The Guidelines indicate that an assessment of a pesticide or any other chemical includes the following three parts:

1. Hazard identification/dose-response involves the evaluation of all available experimental animal and human data and the associated doses, routes and durations of exposure to determine if an agent causes reproductive toxicity in that species and under what exposure conditions.

2. Exposure assessment in which the exposed population and conditions of exposure are described.

3. Risk characterization in which parts 1. and 2., are combined to estimate some measure of the risk of reproductive toxicity.

The Peer Review Document should fulfill part 1. The document should be formatted in a manner similar to that of the material submitted to the Carcinogenicity Peer Review committee. The document should follow the outline given below.

Suggested Outline for Reproductive
Toxicity Peer Review Committee Presentations

I. Introduction

(Includes a brief description of the uses for the chemical and its chemical names, synonyms and structure.)

II. Qualitative Assessment of Relevant Data

A. Rat Study #1

1. Description of maternal toxicity (including data to show dose-response and extent of effects). This should provide the Committee with sufficient information to arrive at conclusions regarding the appropriateness of dose selection.

2. Description of reproductive toxicity (including data to show dose-response and type of reproductive effects). This should provide the Committee with sufficient information to arrive at conclusions.

3. Summary of deficiencies and limitations of the study.

B. Rat Study #n (if available)

1. (Same as A. 1., above)
2. (Same as A. 2., above)
3. (Same as A. 3., above)

C. Other Species Study #n (same as A. and B., above) (if available)

III. Other Data

A. Developmental Studies

B. Subchronic and Chronic Toxicity Data

C. Mutagenicity Studies

D. Metabolism/pharmacokinetics/physico-chemical Data

E. Structure-Activity Relationships (particular concern should be placed on estrogenic/ antiestrogenic compounds)

IV. Strength of the Evidence

A. Strength of the Evidence

1. the quality of the data,
2. the resolving power of the studies,
3. the number and types of endpoints examined,
4. the relevance of route and timing of exposure,
5. the appropriateness of dose selection,
6. the reproducibility of the effects,
7. the number of species examined,
8. pharmacokinetic data,
9. structure-activity relationships, and
10. other factors

B. Questions to the Committee

A minimum of two appendices should follow the document outlined above. They should contain the DER's for the reproductive toxicity and developmental toxicity studies (Appendix 1) and the Toxicology "One-Liners" (Appendix 2). Additional appendices may be also needed (e.g., historical control data).

B. Hazard/exposure issues

The following section is intended to provide perspective for selected issues which often arise in performing a reproductive toxicity risk assessment and to offer guidance in addressing these issues. It is not intended to encompass all situations but to provide the reader with some concrete examples of the issues discussed.

1. Data generated from oral (gavage), dietary, dermal or inhalation exposures

Reproductive toxicity data are primarily generated using dietary exposures. Because pesticides are often found as residues on raw agricultural foods and related feeds, studies by this route are generally appropriate in the assessment of tolerances. Oral intubation (gavage) allows precise measurement of dosage and is

sometimes performed with volatile chemicals. Dietary risk assessments are conducted with the oral NOEL for reproductive toxicity without concerns for route-to-route extrapolation. Dietary exposure may result in a different pattern of developmental effects from that observed with oral intubation (e.g. Giovanni et al., 1986). Studies of pesticides which are used as gases may be performed via inhalation. For more information on some general requirements for the inhalation route the reader is directed to the OPP SEP for inhalation studies.

Despite the fact that most exposure of applicators, mixer/loaders, bystanders and harvesters is by the dermal route, testing by this route is rare. A dermal absorption study is often submitted to clarify the potential dermal penetration to allow better estimates of the margin-of-exposure.

2. Post-natal data: neurotoxicity and other special studies

Many pesticides, including those with neurotoxicity, will have to be evaluated for their potential effect upon the structure and functioning of the nervous system in offspring exposed during pregnancy and lactation (Pesticide Assessment Guidelines, Subdivision F, Hazard Evaluation: Human and Domestic Animals, Series 81, 82, and 83, Neurotoxicity, 1991). The need for postnatal evaluation should be expanded to include other systems than the nervous system when the compound has a hormonal action such as those discussed under the reproductive studies section (B. 11.) below. In that case, a special protocol must be developed to evaluate endocrine-active agents.

In developmental neurotoxicity studies, the test substance is administered in the female rat from gestation day 6 through day 10 of lactation. Dosing (usually oral) is not performed on the day of parturition in animals who have not completely delivered their offspring. The neurotoxicity evaluation includes observations to

detect neurologic and behavioral abnormalities, determination of motor activity, response to auditory startle, assessment of learning, neuropathological evaluation, and brain weights. This type of study may either be separate from an adult neurotoxicity study or be part of a multigeneration reproduction study.

3. Mechanism(s) of action

Knowledge of the site and mechanism of action of observed reproductive/developmental toxicity in animals studies can either diminish or increase our concern for the human population. If the mechanism of toxicity is or is likely to be operative in human physiology (e.g., same receptor site, same metabolic activation route), then much greater weight should be placed on the experimental findings. General mechanisms of reproductive toxicity include direct acting toxicants (similar structures to endogenous chemicals, chemically reactive agents) and indirect acting compounds (altering hormonal control of the reproductive system, requiring metabolic activation, disrupting homeostasis) (see Table 7 below for chemical examples) (Mattison, 1987).

Table 7. Mechanisms of action of reproductive toxicants[a]

| Mechanism | Compound |
|---|---|
| Direct-acting Reproductive Toxicants | |
| Structural similarity | Steroid hormones<br>Cimetidine<br>Diethylstilbestrol<br>Azathioprine<br>6-Mercaptopurine<br>Halogenated polycyclic hydrocarbons |
| Chemical reactivity | Alkylating agents<br>Cadmium<br>Boron<br>Lead<br>Mercury |
| Indirect-acting Reproductive Toxicants | |

| | |
|---|---|
| Metabolic activation | Ethanol<br>Chlorcyclizine<br>Dibromochloropropane<br>Polycyclic aromatic hydrocarbons<br>Cyclophosphamide<br>Ethylene dibromide |
| Disrupted homeostasis | Salicylazosuphapyridine<br>Halogenated polycyclic hydrocarbons<br>Anticonvulsants<br>Ethanol |

a Taken from Table 5 of Mattison (1987) as modified from Mattison (1983, 1984).

4. Mutagenicity, carcinogenicity and cytotoxicity

Studies show that oocytes and sperm possess DNA repair capabilities (Lee and Dixc-, 1978; Pederson and Mangia, 1978; Lee, 1983 in Mattison, 1987). In general, agents which are mutagenic/genotoxic may have a range of effects upon germ cells quite similar to that observed in somatic cells including 1) cell death, 2) incorporation and repair of mutations; or 3) incorporation and expression of mutations (Mattison, 1987). Compounds which induce the latter forms of toxicity to germ cells may result in dominant lethality. Paternal induced developmental toxicity has also been associated with certain compounds e.g. lead, cyclophosphimide. A recent conference explored the evidence for male-mediated developmental and reproductive toxicity (International Conference on Male-Mediated Developmental Toxicity, September 16-19, 1992). These effects have been suggested to be induced through a variety of mechanisms including 1) genetically-heritable alterations, 2) epigenetic alterations such as disruption of DNA methylation patterns of cytosine residues which may alter the normal developmental program of male germ cell differentiation and subsequently the pregnancy outcome (J. Trasler, 1992), 3) microinjection of toxicant via the ejaculate (Silbergeld, 1992), 4) direct effects upon the epididymis which alters maturation of the sperm (Robaire, 1992). As the basis for these male-mediated effects become more clearly differentiated, such data will require

reconsideration of our overall approach to testing for reproductive toxicity.

The fact that a compound is a carcinogen does not necessarily mean that the same chemical will be a reproductive/developmental toxicant. However, diethylstilbestrol stimulates estrogen receptor-containing tissues tissue and increases the risk of vaginal adenosis (75%), vaginal adenocarcinoma (0.01%) and anomalies in males (25%) (Ruddon, 1990). Other chemicals such as ethylene dibromide and dibromochloropropane are directly genotoxic and induce both reproductive and carcinogenic effects through this same mechanism.

Direct cytotoxic agents (e.g., triphenyl tin hydroxide, dinoseb) are more likely to produce reproductive/developmental toxicity depending upon whether or not a cytotoxic threshold dose is achieved (Jelovsek et al., 1990). While it can be argued that such an effect is due to a general systemic effect rather than a specific reproductive mechanism, rapidly dividing germ cells in the conceptus may result in reproductive toxicity at dose levels lower than those at which general systemic toxicity is observed. The focus should be on the resultant reproductive/developmental effect and the relevant (potential) human exposure.

5. Reproductive data: endocrine-active compounds

Reproductive data (single or multi-generation) can be extremely useful in confirming findings of developmental toxicity from a given chemical or sometimes in identifying developmental toxicity which may not otherwise be observed after dosing during the period of major organogenesis. Continued exposure during the entire gestational period potentially allows the entire period of fetal development to be affected. When such effects are noted, it is reasonable to evaluate the necessity of additional developmental studies incorporating specific postnatal parameters.

One significant example of this relates to the differentiation in the fetus of the muellerian ducts (embryonic tubes from which develop the oviducts, uterus and vagina) and wolffian ducts (embryonic tubes from which develop the ductus deferens, ductus epididymis, seminal vesicle, ejaculatory duct, ureter and pelvis of kidney) into the female and male reproduction organs and accessory tissues during the latter period of organogenesis (Taber, 1970; Williams, 1974). These embryonic tissues are significantly affected by the endocrine environment (androgen) which directs the subsequent development of the fetus into male or female offspring. Specifically, the stimulation of the primitive wolffian ducts to differentiate appears to require testosterone but fails to effect involution of the muellerian structures.

Compounds which have androgenic/anti-androgenic activity may affect normal sexual differentiation. Procymidone [N-(3,5-dichlorophenyl)-1,2-dimethylcyclopropane-1,2-dicarboximide] is a fungicide which has recently been shown to be negative when tested for developmental toxicity in two species (rat, rabbit) at quite high dose levels (up to 300 to 1000 m/kg/day) (EPA Peer Review of Procymidone, October 31, 1990). However, in a two-generation reproduction study, dietary levels of 750 ppm produced reproductive/developmental toxicity including abnormalities of external genitalia (reduced anogenital distance and hypospadias). Since procymidone has been shown to have a low affinity for androgen receptors in prostate cytosol (0.07% of dihydro-testosterone), it is likely that the developmental effects noted were mediated by a disturbance in endocrine function. Vinclozolin [3-(3,5-dichlorophenyl)-5-methyl-5-vinyl-1,3-oxazolidin-2,4-dion], a structural analog of procymidone, produces a similar pattern of developmental effects in male rats when administered either dermally or orally (gavage) in Segment II-type studies during gestation days 6-19 but not when administered on days 6-15 of gestation.

6. Pharmacokinetic/physiologic considerations

Pharmacokinetic studies/modeling are assuming a much greater risk assessment role within the EPA by linking exposure with developmental effects observed in the fetus (animal, humans), and being utilized to properly design developmental toxicity studies. Tables 8 and 9 are presented below with information useful to pharmacokinetic considerations. The reader is directed to the recent symposium and reviews on pharmacokinetics in developmental and reproductive toxicity for further reading on approaches being taken (Kavlock, 1991; Nau and Scott, 1987; Ribeiro and Faustman, 1990).

Table 8. Body weights, surface areas, and conversion factors of dosing from mg/kg into $mg/m^2$ [a,b]

| Species | Body Wt. (kg) | Surface Area ($m^2$) | Conversion Factor | Dose Equivalent (kg)[c] |
|---|---|---|---|---|
| Mouse | 0.02 | 0.0066 | 3.0 | 12.0 |
| Rat | 0.15 | 0.025 | 5.9 | 6.0 |
| Dog | 8 | 0.40 | 20 | 1.7 |
| Monkey | 3 | 0.24 | 12 | 3.0 |
| Human | | | | |
| Child | 20 | 0.80 | 25 | 1.5 |
| Adult | 60 | 1.60 | 37 | 1.0 |

a Nau and Scott (1987), p.95.

b To convert a $mg/kg_2$ dose in a given species into an equivalent $mg/m^2$ dose, the dose is multiplied by the conversion factor.

c Dose equivalent for the adult human is set as 1.0.

Table 9. Physiological characteristics of various species relevant for pharmacokinetics[a]

| Physiological Characteristics | Species | | | | | | |
|---|---|---|---|---|---|---|---|
| | Mouse | Rat | Guinea Pig | Rabbit | Dog | Monkey | Man |
| Bile flow (ml/kg x day) | 100 | 90 | 230 | 120 | 12 | 25 | 5 |
| Urine flow (ml/kg x day) | 50 | 200 | --- | 60 | 30 | 75 | 20 |
| Cardiac output (ml/min x kg) | 300 | 200 | --- | 150 | 100 | 80-300 | 60-100 |
| Hepatic blood flow (L/min) | 0.003 | 0.017 | 0.021 | 0.12 | 0.68 | 0.25 | 1.8 |
| (ml/min x kg) | 120 | 100 | --- | 50 | 25 | 25 | 25-30 |
| Liver weight (% of b.wt.) | 5.1 | 4.0 | 4.6 | 4.8 | 2.9 | 3.3 | 2.4 |
| Renal blood flow (ml/min x kg) | 30 | --- | --- | --- | 22 | 25 | 17 |
| Kidney clearance (ml/min x kg) | 5 | --- | --- | --- | 3.2 | 3 | 1.3 |

a Nau and Scott (1987), p. 95.

a. Comparative Pharmacokinetics

Significant pharmacokinetic differences between humans and animals are evident (Nau, 1991). It is important to note that the half life of xenobiotics are often an order of magnitude shorter in experimental animals than in humans. During conventional developmental toxicity studies, steep concentration-time peaks are often produced due to rapid absorption and elimination; these high peaks rapidly fall to low levels. However, target sites in the human may be exposed to the toxicant for longer periods of time due to the longer half-lives. Furthermore, the first-pass effect (rapid metabolism in the liver due to the direct transport of orally administered chemicals from the gut via the portal vein) is often much more extensive in animals than in humans. Human maternal plasma protein binding is often more extensive than in the plasma of experimental animals. Finally, the duration of the sensitivity of particular developmental processes are often several-fold shorter in experimental animals than in man. Thus, in experimental studies a multiple-dosing regimen during a defined period may be more useful than the conventional once-daily administration regimen, where the time of maximal sensitivity may be missed.

Reproductive toxicity (single, multigeneration) tests are preponderantly dietary in nature and therefore, the peak plasma concentrations are usually lower and of longer duration than after an oral bolus of the same compound. This is due to frequent food consumption in rodents and the slower absorption of compound from the gastrointestinal tract due to the presence of the food would yield a more persistent presence in the blood. However, more extensive metabolism and/or binding to plasma proteins due to the lower plasma concentrations may also be observed after the dietary administration. For significant inhalation or dermal exposures, where the likelihood of much higher peak plasma concentrations is present, the standard reproduction test may be inadequate to address the reproductive potential of a pesticide and special

studies may be required.

b. Physiological alterations during pregnancy

During pregnancy physiological changes in several systems can alter the pharmacokinetics in both the mother and fetus (see Table 10, from Mattison, 1991). These physiological alterations are required for successful pregnancy and lactation and result from maternal homeostatic mechanisms to deliver essential nutrients to the fetus and remove heat, carbon dioxide, and waste products from the fetus. These alterations are species dependent (e.g., cardiac output is increased 50% in humans during pregnancy, 20% in rabbits) and may involve different physiological strategies.

Table 10. Physiological changes during pregnancy[a]

| Parameter | Change |
|---|---|
| Absorption | |
| Gastric emptying time | Increased |
| Intestinal motility | Decreased |
| Pulmonary function | Increased |
| Cardiac output | Increased |
| Blood flow to skin | Increased |
| Distribution | |
| Plasma volume | Increased |
| Total body water | Increased |
| Plasma protein | Decreased |
| Body fat | Increased |
| Metabolism | |
| Hepatic metabolism | ± |
| Extrahepatic metabolism | ± |
| Plasma proteins | Decreased |
| Excretion | |
| Renal blood flow | Increased |
| Glomerular filtration rate | Increased |
| Pulmonary function | Increased |
| Plasma proteins | Decreased |

a Taken from Mattison et al., 1991.

Two major alterations in xenobiotic elimination are renal and extrarenal elimination pathways. These pathways should be kept in

mind in reviewing the pharmacokinetic/metabolic nature of the agent under consideration.

i. Renal mechanisms

Renal function probably undergoes the greatest physiological changes during pregnancy (Krauer, 1987). Based on the fact that renal plasma flow and glomerular filtration almost double in humans during pregnancy (Davison and Hytten, 1974; as cited in Krauer, 1987), it may be inferred that for xenobiotics which are eliminated predominantly in the urine and are not highly protein bound, plasma concentration is generally lowered, half-life decreased and clearance increased in parallel to the increased renal function (see Table 11 for examples of changes in kinetic parameters).

Table 11. Kinetic parameters of xenobiotics with predominantly renal elimination ($Q_o \leq 0.3$)[a]

| Drug | $Q_o$ | Protein binding (%) | Changes in kinetic parameters | | | |
|---|---|---|---|---|---|---|
| | | | $C_p$ | $V_d$ | $t_{1/2}$ | $Cl_{tot}$ |
| Ampicillin | 0.1 | 15-29 | ↓ | ↑ | ↓ | ↑ |
| Cephacetrile | 0.04 | 23-26 | ↓ | | ↓ | |
| Cephalexin | 0.04 | 15 | ↓ | | | |
| Cephazolin | 0.06 | 84 | ↓ | | ↓ | |
| Cefuroxime | 0.07 | 40 | ↓ | ↑ | ↓ | ↑ |
| Digoxin | 0.3 | 20-40 | ↓ | | | ↑ |
| Kanamycin | 0.03 | 0.3 | ↓ | | | |
| Lithium | 0.02 | 0 | ↓ | | | ↑ |
| Sotalol | 0.1 | 54 | ↓ | | | ↑ |

a From Krauer, 1987.
$Q_o$ = the extrarenal dose fraction, which indicates the percentage of the absorbed dose not excreted unchanged in the urine.
$C_p$ = plasma concentration; $V_d$ = volume of distribution; $t_{1/2}$ = half life; $Cl_{tot}$ = total clearance

ii. Extrarenal elimination (liver)

The most important extrarenal elimination occurs in the liver and is dependent on hepatic blood flow, the capability to metabolize drugs (intrinsic hepatic clearance) and the extent of binding in plasma (Krauer, 1987). Xenobiotic disposition varies and can be assessed from the magnitude of the hepatic extraction ratio. In human pregnancy and labor, intrinsic hepatic metabolism may be altered. Changes in xenobiotic protein binding have been shown to be quite significant during pregnancy (refs. 22-27, Krauer, 1987). Changes in plasma xenobiotic free fraction will only be relevant for those chemicals extensively (>85 to 90%) bound. However, from studies with many different drugs (chlorazepate, etidocaine, labetalol, meperidine, metronidazole, oxazepam, phenobarbitone, phenytoin, propanolol, thiopental, valproate, caffeine, diazepam, metoprolol) it is apparent that no general rule can be derived regarding agents primarily eliminated via the hepatic route.

7. Structure-activity relationships

Structure-activity relationships have only been studied to a limited extent for reproductive/developmental toxicants. Table ? is a presentation of known or potential human reproductive toxicants which have been associated with certain sites of action within the reproductive process. This list, while not comprehensive, may be useful to identify potential structural analogs for chemical agents under consideration for Peer Review. A more complete listing of inferred reproductive toxicants based upon animal data is given in Hayes (1982). Certain agents with known actions such as hormonal activity, alkylating ability, CNS/peripheral nervous system activity or which are highly lipophilic in nature should be suspect due to the obvious susceptibility of the reproductive process to perturbation by the actions of such agents.

8. Human data

In the area of pesticide toxicity, human data is primarily available for establishing exposure rather than establishing reproductive hazard. Such data may result from the results of acute poisoning cases or biological monitoring (urine, blood, dermal patches) of field workers or mixer-applicators, derived from state, county or registrant-sponsored studies. Exposure estimates may also be obtained by surrogate analyses.

Exposure assessment is framed differently for developmental risk as opposed to reproductive risk since potentially even single acute/short-term exposures may result in a developmental insult whereas reproductive toxicity is generally associated with subchronic to chronic dietary exposure. Compounds, such as dinoseb and TPTH, which are very acutely toxic or direct acting mutagens (e.g., alkylating agents), may be of particular concern. In assessing the need for Reproductive Toxicity Peer Review, the potential exposure estimates should be factored in early in order to determine if a significant risk is posed [see EPA Guidelines for Assessment of Reproductive Risk (1992) for a fuller discussion of potential human exposure assessments].

9. Measurements of additional endpoints not currently required in FIFRA Guidelines

There is an ongoing reexamination of the adequacy of the reproductive toxicity testing protocols both at the Program and Agency level to evaluate potential male and female reproductive toxicity (see EPA Guidelines for Assessment of Reproductive Risk, 1992). A summary of potential endpoints/markers for incorporation possibly in 1) experimental animal subchronic, reproductive or special studies or 2) direct human assessments is presented below. It is likely that at least some of these endpoints will find their way into future evaluations of reproductive testing and will allow

a more comprehensive assessment of the ability of a chemical agent to disturb reproductive physiology.

a. Male endpoints

Recent laboratory and epidemiologic data suggest that chemical-induced effects upon the offspring, e.g., birth defects, cancer, or death of the conceptus, may be mediated, in some exposures, through an effect upon the male alone (International Conference on Male-Mediated Developmental Toxicity, September 16-19, 1992, Pittsburgh, PA). These effects have been suggested to be mediated by a variety of mechanisms including 1. genetically-heritable alterations (mutations) (Russel, 1992), 2. epigenetic alterations such as disruption of DNA methylation patterns of cytosine residues which may alter the normal developmental program of male germ cell differentiation and subsequently the pregnancy outcome (Trasler, 1992), 3. microinjection of toxicant via the ejaculate (Silbergeld, 1992; presentation), and 4. direct effects upon the epididymis which alters maturation of sperm (Robaire, 1992; presentation). As the basis for these male-mediated effects become more clearly differentiated, such data will require reconsideration of the overall approach to testing reproductive toxicity. For example, if a particular cell type in spermatogenesis is known to be sensitive to the test compound, it may not be necessary to expose males to the chemical for a full ten week period prior to cohabitation.

The biological markers presented in Tables 12 A and B are for assessment of physiological or genetic damage in human males but generally have their counterpart in animal testing.

Table 12A. Male endpoints - physiologic[a]

| Tissue or data | Markers of |
|---|---|
| Testis (or biopsy) | Histopathology |
| Seminal sperm | Sperm number<br>Structure<br>Motility<br>Double F bodies<br>Viability<br>Agglutination<br>Penetration and egg interaction<br>  Cervical mucus<br>  Hamster eggs<br>  Nonliving human eggs<br>Internal and surface domains<br>Chromatin structure |
| Other seminal parameters | Physical characteristics<br>Immature germ cells<br>Non-germ cells<br>Chemical composition<br>  Normal and xenobiotic constituents<br>Sertoli cell, Leydig cell, and accessory gland function |
| Blood | Hormone levels |
| Survey and medical records | Fertility status<br>  Standardized fertility ratio<br>  Time to conception |
| Maternal urine | Indicators of early pregnancy |

a From Table 7-1, Biologic markers of physiologic damage to human male reproduction, reviewed in Chapter 7 (NRC, 1989).

Table 12B. Male endpoints - Genetic

| Tissue | Marker |
|---|---|
| Testis | Cytogenetic analyses of cells in mitosis, meiosis I, and meiosis II |
| Semen<br>Sperm | <br>Sperm cytogenetics<br>Sperm DNA and protein adduction<br>Gene mutations in sperm<br>Sperm aneuploidy |
| Immature germ cells | Spermatid micronuclei<br>Cytogenetics of ejaculated meiotic I cells |
| Questionnaire and medical records | Sex ratio<br>Spontaneous abortion<br>Offspring cancer<br>Sentinel phenotypes |
| Offspring tissue | Cytogenetics<br>DNA sequencing<br>Protein mutations<br>Restriction-length polymorphism of DNS<br>RNAase digestion<br>Subtractive hybridization of DNA<br>Denaturing gel electrophoresis of DNA<br>Pulse-field electrophoresis of DNA |
| Maternal urine | Detection of early fetal loss |
| Somatic cell surrogates<br>In white blood cells<br>In red blood cells | <br>HGPRT mutations<br>Hemoglobin mutations<br>Glycophorin A mutations |

a From Table 9-1, Potential markers of genetic damage and heritable mutations in the male germline, reviewed in Chapter 9 (NRC, 1989).

b. Female endpoints

The biological markers presented in Table 13 are for human female assessment of physiological or genetic damage but generally have a counterpart in animal testing.

Table 13. Female endpoints[a]

| Site | Marker |
|---|---|
| Exposure: Chemical analyses for toxicants or metabolites, or mutagenic analysis of body fluids | Blood, urine, saliva<br>Tissues<br>Intact<br>Cytologic specimens<br>Fluids<br>Cerebrospinal fluid<br>Follicular fluid, amniotic fluid<br>Placental tissue.<br>Peritoneal fluid |
| Genotoxic - DNA adducts (chemical specific, generic | Oocytes, ovarian tissue<br>Placental tissue<br>Fetal tissues<br>Maternal serum<br>Fetal serum<br>Unscheduled DNA synthesis<br>Maternal lymphocytes<br>Fetal lymphocytes<br>SCE (sister-chromatid exchange)<br>Maternal lymphocytes<br>Fetal cells<br>Chromosomal aberrations<br>Maternal serum<br>Abortus tissue<br>Chorionic villi<br>Amniotic cells<br>Fetal serum<br>Micronuclei<br>Maternal blood<br>Vaginal/cervical cells<br>Fetal liver cells<br>Fetal lymphocytes<br>Specific-locus mutations |

Table 13. Female endpoints[a]

| Site | Marker |
|---|---|
| Development/aging | Onset of puberty<br>Breast bud development<br>Blood<br>Melatonin<br>DHEA-S<br>Gonadotropin (pulsatile)<br>Age of first menstrual bleeding<br>Hormones: estrogens, inhibin, LH, FSH, androgens<br>Age of breast development<br>Sexual behavior<br>Neurotransmitter in CSF<br>Menstrual cycle length<br>Ovarian-oocyte stock<br>Ultrasound for ovarian size<br>IVF<br>Biopsy<br>MRI<br>Periodic ultrasound to monitor follicular development<br>Inhibin<br>Premenopausal hormonal status (estrogens, gonadotropins, inhibin, LH, FSH)<br>CNS reproductive senescence |
| Menstrual function | Cycle frequency and characteristics<br>Detection of corpus luteum<br>Follicular development (ultrasound)<br>Basal body temperature<br>Thermometer<br>Improved, self-recording electronic thermometer<br>Cervical mucus<br>Sexual behavior<br>Vaginal cytology<br>Biophysical measurements of vaginal secretions<br>Endometrial histology<br>Endocrinology: gonadotropins, steroids, ovulatory hormones<br>In vitro assays<br>Pituitary cells (from cadavers)<br>Granulosa cells<br>Luteal-specific proteins, endometrial cell cultures<br>Mucus production, endocervical cells |
| Fertilization, implantation, and loss | hCG (human chorionic gonadotropin)<br>EPF (early pregnancy factor)<br>PEP (progestin-associated endometrial protein) |

a From Table 16-1, Status of current and potential markers in female reproductive toxicology, Chapter 16 (NRC, 1989).

C. Risk characterization

Risk characterization is the culmination of hazard assessment/dose-response and exposure assessment.

1. Reference Dose vs Margins-of-Exposure (Margins-of-Safety)

a. Use of the reproductive No-Observed-Effect-Level (NOEL)

Current practice within the Agency is to characterize developmental risk by using margins-of-exposure (MOE) (formerly known as margins-of-safety) or uncertainty factors. The MOE is a direct comparison (ratio) between the appropriate No-Observed-Effect Level (NOEL) and the estimated human exposure. The uncertainty factor approach divides by uncertainty factors which generally include a 10-fold factor for interspecies variation and a 10-fold factor for intraspecies variation. Approaches for the generation of Benchmark dose levels, developed from models which utilize data at all dose levels, are available for developmental toxicity studies and may be applicable to reproduction studies in the future.

NOELs from reproduction studies are considered along with systemic toxicity NOELs derived from other data sets (e.g. subchronic, chronic studies in dogs and rodents) in selecting the appropriate study for setting the reference dose (RfD) for a pesticide. Determining what constitutes a selective reproductive effect as opposed to a general systemic effect is a matter of careful analysis of the parental and offspring effects and may not be easily resolvable. It is important to compare the resultant NOELs observed in reproduction studies against other long-term studies to determine if the pregnant animal is more sensitive than the non-pregnant female.

The NOEL from the most sensitive species tested (where

multiple tests are available) is generally used for reproductive toxicity risk assessment purposes due to the great difficulty in determining the most relevant species from which to extrapolate to humans.

b. Subchronic/chronic vs "short-term" exposures

In most instances, exposure assessment in OPP is framed differently for reproductive risk as opposed to developmental risk. This is due to the potential that even single acute/short-term exposures may result in a developmental insult whereas reproductive toxicity studies are generally associated with subchronic to chronic dietary exposure. It is recognized that this approach is an artifact of the testing procedures. Thus, for reproductive risk assessment, the use of NOELs from a reproduction test would not be generally compared against acute/repeated exposure situations such as those observed with mixer-loader-applicators. Rather, the comparison would be between the subchronic/chronic average daily exposure in the test species as compared to the potential or observed subchronic/chronic exposure in the human population of concern.

There are many reasons, related to the physiology of the reproductive system, the presence of highly sensitive individuals within an exposed population, and the nature of the individual chemical agent, which support the possibility that a short-term exposure (acute, high exposure; repeated moderate exposure; acute exposure to a potent reproductive toxicant) may be sufficient to produce reproductive toxicity of either a reversible or irreversible nature. For example, destruction of the Sertoli cells or spermatogonia in the testes or oocytes in the ovaries is an irreversible phenomenon which may occur from a single exposure and may permanently affect the reproductive capacity of the exposed individual. After ovulation, single exposures to compounds such as carbendizim also alter the fertilizability of the ova (Darney,

1990). Such an effect could cause significant delays in the ability of an individual to conceive, particularly one with low fertility potential. Determination that a pesticide has such acute reproductive toxicity potential must be done on a case-by-case basis.

c. Forms of exposure: dietary, occupational, drinking water

Reproductive toxicity risk assessment should include dietary and worker exposure, as well as other forms of exposure such as drinking water or home use. Worker exposure estimates are the responsibility of the Occupational and Residential Exposure Branch (OREB). OREB estimates of exposure are on a daily basis and quantified for each route of exposure.

It is the responsibility of the OPP scientists to determine the rate of dermal absorption. If available, pharmacokinetic data such as peak plasma concentrations, area under the curve of the test material, and/or metabolites should be compared when dosing is by different routes. Metabolism data should also be considered, since this may vary with route of exposure. In the absence of dermal absorption data, a 100% rate of absorption is generally assumed.

Dietary risks should be assessed using the Dietary Residue Exposure System (DRES). This system compares NOELs to the predicted dietary exposure for the appropriate subgroups, e.g., adult males, infants, and children, and for the period of exposure (acute or chronic) which is most appropriate for the form of toxicity which is the basis for the NOEL.

Drinking water risks may also be of concern and are assessed in a manner similar to dietary risk. Determination of whether or not a pesticide has a potential for groundwater or surface water contamination is the responsibility of the Environmental Fate and

Effects Division (EFED). If actual contamination exists, the most relevant contamination levels must be selected in consultation with the EFED. The National Academy of Sciences has recommended that, for risk assessment purposes, it be assumed that the average adult consumes two liters of water per day (Office of Water/US EPA and NAS, 1977). The estimated daily exposure (mg/kg/day) to a pesticide in drinking water is therefore determined by multiplying the appropriate estimate of the residue level (mg/liter) by two liters and dividing that amount by body weight. A more detailed description of the hazard evaluation of pesticides in drinking water can be found in the book "Drinking Water Health Advisory: Pesticides," US EPA, Lewis Publishers, 1989.

Bibliography

Adler, NT., Resko, FA. and Goy, RW., 1970. The effect of copulatory Behavior on hormonal change in the female rat prior to implantation. Physiol. Behavior., Vol. 5, 1003-1007.

Amann, R.P., 1981. A critical review of methods for evaluation of spermatogenesis from seminal characteristics. J. Androl 2:37-58.

Anver, MR. and Coheil, BJ., 1979. Lesions associated with aging. In "The Laboratory Rat", eds., HJ. Baker, JR. Lindsey, and SH. Weisbroth, Vol. 1, Academic Press.

Baetjer, AM., 1968. Role of environmental temperature and humidity In susceptibility to disease. Arch. Environ. Healths, Vol. 16, 565-570.

Baker, HJ., Lindsey, JR. and Weisbroth, SH., 1979. Housing to control research variables. In "The Laboratory Rat", eds., HJ. Baker, JR. Lindsey and SH. Weisbroth, Vol. 1, Academic Press.

Bellhorn, RW., 1986. Lighting in the animal environment. Lab. Animal Sci., Vol. 10, 440-450.

Berg, BN., 1965. Dietary restriction and reproduction in the rat. J. Nutrition, Vol. 87, 344-348.

Blazak, WF., Ernst, TL. and Stewart, BE., 1985. Potential indicators of reproductive toxicity: testicular sperm production and epididymal sperm number, transit, time, and motility in Fischer 344 rats. Fund. Appl. Tox., Vol. 5, 1097-1103.

Cheeke, PR. and Patton, NM., 1982. In "Rabbit Production", Inter-state Printel-s and Publishers, Inc.

Chester, RV. and Zticker, I., 1970. Influence of male copulatory behavior on sperm transport, pregnancy, and pseudopregnancy in female rats. Physiol. Behavior., Vol. 5, 35-43.

Christian, MS., 1986. A critical review of multigeneration studies. J. Amer. Coll. Tox., Vol. 5, 161-180.

Clegg, D.J. Animal Reproduction and Carcinogenicity Studies in Relation to Human Safety Evaluation in Toxicology and Occupational Medicine ed. Diechman, Elsevier, North Holland, pp. 44-59.

Clemens, G., F. Fort, K. Olberholtzer, I. Lamb. Midwest Teratology Association (MTA) Historical Control Database Survey (HCDS), Phase 1 (P1): Maternal Reproductive (Segment II) Indices. Poster presented at Teratology Society Annual Meeting, 1992.

Collins, T.F.X., 1978. Multigeneration reproduction studies. In "Handbook of Teratology", Vol. 4, eds. JG. Wilson and FC. Fraser, Plenum Press.

Collins, T.F.X., 1978. Reproduction and developmental toxicity guidelines: Review of deliberations by the National Toxicology Advisory Committee's reproductive panel. J. Environ. Pathol. Toxicol., 2:141-47.

Collins, TFX., Keeler, HV., Black, TN. and Ruggles, DI., 1975. Long term effects of dietary Amaranth in rats. I. Effects on reproduction. Toxicology Vol. 3, 115-128.

Degraeve, N., 1981 Carcinogenic, teratogenic, and mutagenic effects of Cadmium Mutation Res. Vol. 86, 115-135.

Dixon, RL. and Hall, JL., 1982. In "Reproductive Toxicology", eds. A. Wallace Hayes, Raven Press, New York.

Fitzhugh, OG., 1968 Reproductive tests. In "Modern trends in Toxicology", eds. E Boyland and R. Goulding, Vol. 1, Butterworth, London.

Foote, 1986, Use of quantitative testicular histology to assess the effect of dibromochloropropane (DBCP) on reproduction in rabbits. Fund App Tox 6:638-647

Frisch, RE., 1978. Population, food intake, and fertility. Science Vol. 199, 22-30.

Goldenthal, E.I., 1966. Guidelines for Safety Evaluation of Drugs for Human Use. Drug Review Branch, Division of Toxicological Evaluation, Bureau of Science, Food and Drug Administration, Washington, DC.

Gray, L.E., Ostby, L., Linder, R., Goldman, J., Rehnberg, G. and Cooper, R., 1990. Carbendizim-induced alterations of reproductive development and function in the rat and hamster. Fund Appl Tox 15: 281-297.

Harness, JE. and Wagner, JE., 1977. In "Biology and medicine of rabbits and rodents", Lea and Fibiger.

Haseman J. and Soares, 1976. The distribution of fetal death in control mice and its implications on statistical tests for dominant lethal effects. Mut Res 41:277-288.

Hastings, JW. and Meneker, AH., 1976. Physiological and biochemical aspects of circadian rhythms. Fed. Proc. Am. Soc. Exp. Biol., Vol. 35 2325-2357.

HED Ad hoc Committee for Atrazine Reproductive Issue. 1992.

Memorandum of September 23, 1992, "Atrazine Two-Generation Study". US EPA.

Holmes, DD., 1985. The mongolian gerbil in biomedical research. Lab. Animal Sci., Vol. 14, 23-38.

Jacoby, RO. and Fox, JG., 1984. Biology and diseases of mice. In "Laboratory Animal Medicine", eds., JG. Fox, BJ. Cohen, and FM. Loew, Academic Press.

Johnson, E.M., 1986. The Scientific Basis for multigeneration Safety Evaluations, J.Am. Col.Tox., 5:197-201.

Khera, K., Grice, H.C. and Clegg, D., 1989. Interpretation and Extrapolation of Reproductive Data to Establish Human Safety Standards, Current Issues in Toxicology. Current Issues in Toxicology, ILSI, Springer-Verlag, New York.

Kohn, DF. and Barthuld, SW., 1984. Biology and diseases of rats. In "Laboratory Animal Medicine", eds., JG. Fox, BJ. Cohen, and FM. Loew, Academic Press.

Lamb and Chapin, 1985.

Merry, BJ. and Holetian, AM., 1979. Onset of puberty and duration of fertility in rats fed a restricted diet. J. Reprod. Fertility Vol. 57, 253-259.

Moore, K.L., 1982. The Developing Human- Clinically oriented embryology. W.B. Saunders, Philadelphia.

Mosher, W.D., Pratt, W.F., 1985. Fecundity and Infertility in the United States, 1965-1982 Demography 22(3):415-430.

Morrissey et al, 1989.

Mulder, JB., 1971. Animal behavior and electromagnetic energy waves. Lab. Animal Sci., Vol. 21, 389-393.

National Research council, 1977. Reproduction and Teratogenicity Tests. In "Principles and Procedures for Evaluating the Toxicity of Household Substances", National Academy of Sciences.

National Research Council, 1978. Nutrient requirements of laboratory Animals 3rd revised ed., Vol. 10, National Academy of Sciences.

Nelson, JF. and Felicio, LS., 1984. Dietary modulations of estrous cyclicity in singly and multiply housed C57BL/6J mice. Lab. Animal Science Vol. 34(2), 173-176.

Newman, L.M., Johnson, E.M., and Cadogan, A.S.A., 1983. Oxygen

consumption and survival prediction in neonatal rats exposed to prenatal hypervitaminosis A. Teratology 28, 219-227.

Newman, L.M., Johnson, E.M., and Roth, J.M., 1984. Lung volume and compliance in neonatal rats. Lab Animals Science 34(4), 371-375.

Palmer, A.K., 1986. A simpler multigeneration study. Poster presented at the 6th International Congress of Pesticide Chemistry, Ottawa.

Palmer, A.K., 1981. Regulatory Requirements for Reproductive Toxicology: Theory and Practice. In: Developmental Toxicology, edited by C.A. Kimmel and J. Buelke-Sam, pp.259-287. Raven Press, New York.

Pakes, SP., Lu, YS. and Meunier, PC., 1984. Factors that complicate animal research. In "Laboratory Animal Medicine", eds., JG. Fox, BJ. Cohen, and FM. Loew, Academic Press.

Robinson, DG., 1979 Endocrinological research. The Gerbil Digest, Vol. 6, 1-4

Schwartz, NB., Dierschke, DJ., McCormack, CE. and Waltz, PW., 1977. Feedback regulations of reproductive cycles in rats,

United States Environmental Protection Agency, 1988. Draft of Proposed Guidelines for the Risk Assessment of Suspect Female Reproductive Toxicants, Office of Research and Development.

United States Environmental Protection Agency, 1987. Workshop on One- vs. Two-Generation Reproductive Effects Studies, Sponsored by the U.S.E.P.A. Risk Assessment Forum, Washington, DC.

United States Environmental Protection Agency, 1986. Guidelines for estimating exposures. Federal Register 51(185):34042-34054

Whorton and Milby, J. Occup. Med, 22, 1980

From Jim Rowe's write-up on male-mediated effects (get the details from him):

International Conference on Male-Mediated Developmental Toxicity, 1992.

Russel, 1992

Trasler, 1992

Silbergeld, 1992

Robaire, 1992

Table . Some parameters of reproduction

| Event | Human | Rhesus | Rat | Mouse | Rabbit | Beagle | Guinea Pig | Hamster | Ferret |
|---|---|---|---|---|---|---|---|---|---|
| Male breeding age | 14 yr | 3-6 yr | 8-12 wk | 8-10 wk | 6-7 mo | 10-14 mo | 3-5 mo | | 7-10 mo |
| Female breeding age | 13 yr | 2-5 yr | 8-12 wk | 6-8 wk | 5-6 mo | 10-14 mo | 3-5 mo | | 7-10 mo |
| Type of estrous cycle | | | a | a | a | Biannual | a | a | b |
| Length of cycle | 28 d | 28 d | 4-5 d | 4-5 days | c | 6 mo | | | |
| Duration of estrus | | | 10-20 hr | 10-20 hr | c | | 6-15 days | 12-20 hr | |
| Time of ovulation (hours after onset of estrus) | | | 8-11 | 2-3 | 10 p.c.[d] | | 10 | 8-12 | |
| Gestation length (days) | 250-267 | 150-174 | 20-22 | 19-21 | 28-35 | 58-65 | 63 | 16-17 | 39-43 |
| Litter size (#) | 1 | 1 | 8-12 | 4-12 | 7-9 | 1-15 | 1-6 | 6-9 | 8-12 |
| Birth weight (g) | | | 5-6 | 1-3 | 30-70 | | 85-90 | 2-3 | |
| Weaning age (day) | | | 21 | 21 | 50 | | 20 | 21 | |
| Weaning weight (g) | | | 40-50 | 10-12 | 800-1500 | | 180-240 | 35-40 | |
| Age at puberty (mo) | | 18-48 | 2-3 | 1.5-2 | 5-7 | 9-14 | 3-5 | 2 | 5-7 |
| Weight at puberty (g) | | 3500-6000 | 200-300 | 20-35 | 3500-4000 | Variable | 500-550 | 95-120 | |
| Duration of ability to reproduce (years) | | 30-35 | 1 | 1-1.5 | 1-3 | 6-14 | 4-5 | 1 | |

a Polyestrous
b Monoestrous
c Continuous
d Postcoital
Adapted from Hafez, 1970 and McLain, 1985.

Table . Fertilizable life of the egg, stages of egg development, and implantation in various species

| Animal | Fertilizable Life of Egg (hr) | Day Found | | | | |
|---|---|---|---|---|---|---|
| | | 2-Cell | 4-Cell | 16-Cell | Blastocyst | Implantation |
| Human | 12-24 | 2 | early 3 | late 3 | 4-5 | 6 |
| Monkey | 23 | 2 | early 3 | late 3 | 5 | 9 |
| Rat | 12 | 2 and 3 | late 3 | late 4 | 5 | late 5 |
| Mouse[a] | 15 | 2 | early 3 | late 3 | early 4 | early 5 |
| Rabbit[a] | 6-8 | 2 | late 2 | early 3 | late 3 | 7 |
| Dog | 24 | | | | | |
| Guinea Pig | 20 | 2 | 4 | 5 | late 5 | 6 |
| Hamster | 5 | 2 | 3 | early 4 | 4 | 5 |
| Ferret | | 3 | 3 | 4-5 | 6-7 | 11-12 |
| Gerbil | | 2 | 3 | late 4 | 5 | 6 |
| Mink[a] | | 3 | 4 | 5-6 | 6-7 | delayed |
| Oppossum[a] | | 3 | 3 | 4 | early 5 | 6 |

a Species with induced ovulation; all others have spontaneous ovulation.
From Hafez, 1970 and Moore, 1982.

Table . Comparison of prepubertal ovarian development in some laboratory animals

| Stage of Egg Development | Age (days) | | | | |
|---|---|---|---|---|---|
| | Guinea pig | Hamster | Mouse | Rabbit | Rat |
| Early postnatal oogenesis | No | Yes | No | Yes | No |
| Initiation of interstitial development | | Prim.: 18 | Prim.: 11 | Prim.: 60 | Prim.: 11 |
| | Sec.: 21 | Sec.: 28+ | Sec.: 23 | Sec.: 86 | Sec.: 25 |
| First appearance of antral follicles | 21-60 | 26 | 14 | 64-70 | 12 |
| Earliest induced ovulation or follicular stimulation | 65-105 | 27 | 14 | 65-105 | 18 |
| Earliest spontaneous ovulation | 139 | 30 | 35 | 139 | 38-71 |

Prim. = Primary; Sec. = Secondary.
From Hafez, 1970 (compiled from Green and Peppler, 1986).